

No. 24-6136

---

**UNITED STATES COURT OF APPEALS  
FOR THE NINTH CIRCUIT**

---

JANE DOE 1 ET AL.,

*Plaintiffs-Petitioners,*

v.

GITHUB, INC., MICROSOFT CORPORATION, OPENAI, INC., OPENAI, L.P., OPENAI  
OPCo, L.L.C., OPENAI GP, L.L.C., OPENAI STARTUP FUND GP I, L.L.C., OPENAI  
STARTUP FUND I, L.P., OPENAI STARTUP FUND MANAGEMENT, LLC, OPENAI LLC,  
OPENAI GLOBAL, LLC, OAI CORPORATION, OPENAI HOLDINGS, LLC, OPENAI  
HOLDCo LLC, OPENAI STARTUP FUND SPV I, LP, OPENAI STARTUP FUND SPV GP I  
LLC, AND OPENAI INVESTMENT,

*Defendants-Respondents.*

---

From a Certified Order of the United States District Court for the  
Northern District of California, Oakland Division  
Case No. 22-cv-6823 (JST)

---

**OPENAI DEFENDANTS-RESPONDENTS' ANSWER TO PLAINTIFFS'  
PETITION FOR PERMISSION TO APPEAL ORDER CERTIFIED UNDER  
28 U.S.C. § 1292(B)**

---

ALEXANDRA M. AVVOCATO  
MORRISON & FOERSTER LLP  
250 West 55th Street  
New York, NY 10019

AILEEN M. MCGRATH  
JOSEPH C. GRATZ  
TIFFANY CHEUNG  
MORRISON & FOERSTER LLP  
425 Market Street  
San Francisco, CA 94105  
Tel.: (415) 268-6153  
AMcGrath@mofo.com

*Counsel for OpenAI Inc., OpenAI, L.P., OpenAI OpCo, L.L.C., OpenAI GP,  
L.L.C., OpenAI Startup Fund GP I, L.L.C., OpenAI Startup Fund I, L.P.,  
OpenAI Startup Fund Management, LLC, OpenAI LLC, OpenAI Global,  
LLC, OAI Corporation, OpenAI Holdings, LLC, OpenAI HoldCo LLC,  
OpenAI Startup Fund SPV I, LP, and OpenAI Startup Fund SPV GP I LLC*

OCTOBER 17, 2024

---

## **CORPORATE DISCLOSURE STATEMENTS\***

OpenAI Inc. and OpenAI Startup Fund GP I, L.L.C. state that they have no parent company and no publicly held corporation owns 10% or more of their stock.

OpenAI Startup Fund I, L.P. states that it is a subsidiary of OpenAI Startup Fund GP I, L.L.C. and Microsoft Corporation has a financial interest of 10% or more in OpenAI Startup Fund I, LP.

OpenAI GP, L.L.C., OAI Corporation, and OpenAI Holdings, LLC state that they are subsidiaries of OpenAI, Inc. and no publicly held corporation has a financial interest of 10% or more.

OpenAI OpCo, L.L.C. (OpenAI OpCo) states that OpenAI, L.P. recently changed its name to OpenAI OpCo, L.L.C. OpenAI OpCo further states it is a subsidiary of OpenAI Global, LLC and that Microsoft Corporation is a publicly traded corporation that has a financial interest of 10% or more in OpenAI Global, LLC.

OpenAI HoldCo LLC was dissolved in April 2023, and was a subsidiary of OpenAI LLC.

OpenAI LLC states that it is a subsidiary of OpenAI Global, LLC and that Microsoft Corporation is a publicly held corporation that has a financial interest of 10% or more in OpenAI Global, LLC.

OpenAI Global, LLC states that it is a subsidiary of OpenAI Inc. and that Microsoft Corporation is a publicly held corporation that has a financial interest of 10% or more in OpenAI Global, LLC.

OpenAI Startup Fund Management, LLC and OpenAI Startup Fund SPV GP I LLC state that they are subsidiaries of OpenAI Startup Fund GP I, LLC and that no publicly held corporation has a financial interest of 10% or more.

OpenAI Startup Fund SPV I, LP states that it is a subsidiary of OpenAI Startup Fund SPV GP I, LLC and that no publicly held corporation has a financial interest of 10% or more.

---

\* Although OpenAI Investment appears as a party on the Court's docket, it was dismissed in the district court, and is unaffiliated with any of the remaining OpenAI entities.



Dated: October 17, 2024

s/ Aileen M. McGrath  

---

Aileen M. McGrath

## TABLE OF CONTENTS

TABLE OF AUTHORITIES .....	iv
INTRODUCTION .....	1
STATEMENT OF THE CASE.....	3
REASONS THE PETITION SHOULD BE DENIED .....	5
I. ANY QUESTION OF LAW PRESENTED HERE IS NOT CONTROLLING .....	6
II. THE PETITION DOES NOT IDENTIFY A SUBSTANTIAL GROUND FOR A DIFFERENCE IN OPINION .....	10
III. IMMEDIATE APPEAL WOULD NOT MATERIALLY ADVANCE THE LITIGATION .....	16
CONCLUSION .....	22

## TABLE OF AUTHORITIES

	Page(s)
<b>Cases</b>	
<i>ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.</i> , 667 F. Supp. 3d 411 (W.D. Tex. 2023) .....	12, 13
<i>Andersen v. Stability AI Ltd.</i> , No. 23-cv-0201, 2024 WL 3823234 (N.D. Cal. Aug. 12, 2024).....	15
<i>Beijing Meishe Network Tech. Co. v. TikTok Inc.</i> , No. 23-cv-06012, 2024 WL 3522196 (N.D. Cal. July 23, 2024).....	15
<i>Bounce Exchange, Inc. v. Zeus Enterprises, Ltd.</i> , No. 15-cv-3268, 2015 WL 8579023 (S.D.N.Y. Dec. 9, 2015) .....	14
<i>In re Cement Antitrust Litig.</i> , 673 F.2d 1020 (9th Cir. 1981) .....	9, 19
<i>Coopers &amp; Lybrand v. Livesay</i> , 437 U.S. 463 (1978).....	6
<i>Couch v. Telescope Inc.</i> , 611 F.3d 629 (9th Cir. 2010) .....	6, 10, 11, 12, 16
<i>Dolls Kill, Inc. v. Zoetop Bus. Co.</i> , No. 22-cv-1463, 2022 WL 16961477 (C.D. Cal. Aug. 25, 2022).....	11-12
<i>Enter. Tech. Holdings, Inc. v. Noveon Sys., Inc.</i> , No. 05-cv-2236, 2008 WL 11338356 (S.D. Cal. July 29, 2008) .....	15
<i>Fast v. Applebee’s Int’l, Inc.</i> , 638 F.3d 872 (8th Cir. 2011) .....	8
<i>Faulkner Press, L.L.C. v. Class Notes, L.L.C.</i> , 756 F. Supp. 2d 1352 (N.D. Fla. 2010) .....	12



<i>Fischer v. Forrest</i> , 286 F. Supp. 3d 590 (S.D.N.Y. 2018) .....	12
<i>In re Flor</i> , 79 F.3d 281 (2d Cir. 1996) .....	11
<i>Frost-Tsuji Architects v. Highway Inn, Inc.</i> , No. 13-cv-0496, 2014 WL 5798282 (D. Haw. Nov. 7, 2014) .....	11, 16
<i>Garcia v. Johanns</i> , 444 F.3d 625 (D.C. Cir. 2006) .....	21
<i>Matter of Hamilton</i> , 122 F.3d 13 (7th Cir. 1997) .....	10
<i>U.S. ex rel. Huangyan Import &amp; Export Corp. v. Nature's Farm Prods., Inc.</i> , 370 F. Supp. 2d 993 (N.D. Cal. 2005) .....	19
<i>ICTSI Or., Inc. v. Int'l Longshore &amp; Warehouse Union</i> , 22 F.4th 1125 (9th Cir. 2022) .....	15-16
<i>Int'l Soc'y for Krishna Consciousness, Inc. v. Air Canada</i> , 727 F.2d 253 (2d Cir. 1984) .....	21
<i>Joffe v. Google, Inc.</i> , 746 F.3d 920 (9th Cir. 2013) .....	8
<i>Kipp Flores Architects, LLC v. Pradera SFR, LLC</i> , No. 21-cv-00673, 2022 WL 1105751 (W.D. Tex. Apr. 13, 2022) .....	13
<i>Kirk Kara Corp. v. W. Stone &amp; Metal Corp.</i> , No. 20-cv-1931, 2020 WL 5991503 (C.D. Cal. Aug. 14, 2020) .....	11
<i>Koehler v. Bank of Bermuda Ltd.</i> , 101 F.3d 863 (2d Cir. 1996) .....	22
<i>Mamani v. Berzain</i> , 825 F.3d 1304 (11th Cir. 2016) .....	7, 8

<i>O’Neal v. Sideshow, Inc.</i> , 583 F. Supp. 3d 1282 (C.D. Cal. 2022) .....	12
<i>Oracle Int’l Corp. v. Rimini Street, Inc.</i> , No. 14-cv-1699, 2023 WL 4706127 (D. Nev. July 24, 2023).....	15
<i>In re Prudential Lines, Inc.</i> , 59 F.3d 327 (2d Cir. 1995) .....	10
<i>Rabin v. PricewaterhouseCoopers LLP</i> , No. 16-cv-02276, 2017 WL 11662124 (N.D. Cal. Apr. 17, 2017) .....	10
<i>Real World Media LLC v. Daily Caller, Inc.</i> , No. 23-cv-1654, 2024 WL 3835351 (D.D.C. Aug. 14, 2024) .....	14
<i>Richardson Elecs., Ltd. v. Panache Broad. of Pa., Inc.</i> , 202 F.3d 957 (7th Cir. 2000) .....	8
<i>Sec’y, U.S. Dep’t of Lab. v. Preston</i> , 873 F.3d 877 (11th Cir. 2017) .....	8
<i>Slade v. Shearson, Hammill &amp; Co.</i> , 517 F.2d 398 (2d Cir. 1974) .....	21
<i>Software Pricing Partners, LLC v. Geisman</i> , No. 19-cv-0195, 2022 WL 3971292 (W.D.N.C. Aug. 31, 2022).....	13, 15
<i>In re Somberg</i> , 31 F.4th 1006 (6th Cir. 2022) .....	17, 18
<i>Splunk, Inc. v. Cribl, Inc.</i> , 662 F. Supp. 3d 1029 (N.D. Cal. 2023).....	14
<i>Steering Comm. v. United States</i> , 6 F.3d 572 (9th Cir. 1993) .....	6, 7
<i>Sterk v. Redbox Automated Retail, LLC</i> , 672 F.3d 535 (7th Cir. 2012) .....	19

<i>S.B.L. ex rel. T.B. v. Evans</i> , 80 F.3d 307 (8th Cir. 1996) .....	21
<i>United States Rubber Co. v. Wright</i> , 359 F.2d 784 (9th Cir. 1966) .....	10
<i>Wood v. GCC Bend, LLC</i> , 422 F.3d 873 (9th Cir. 2005) .....	18
<b>Statutes</b>	
17 U.S.C. § 1202(b) .....	16
17 U.S.C. § 1202(b)(1).....	3
17 U.S.C. § 1202(b)(3).....	3
28 U.S.C. § 1292(b) .....	1, 6, 10
<b>Other Authorities</b>	
16 Fed. Prac. & Proc. Juris. § 3930 (3d ed.).....	9



## INTRODUCTION

The district court dismissed plaintiffs’ Digital Millennium Copyright Act (DMCA) claims based on its fact-specific determination that plaintiffs failed to plausibly allege that defendants’ artificial-intelligence tool would remove copyright management information (CMI) from identical copies of plaintiffs’ works. That dismissal came after plaintiffs filed serial unsuccessful amended complaints in an attempt to cure those deficiencies. Plaintiffs now try again in this Court, asking it to circumvent ordinary appellate procedure and immediately review a purportedly pure question of statutory interpretation: whether the DMCA requires identity. But plaintiffs do not demonstrate that this case presents the extraordinary circumstances warranting interlocutory appeal under 28 U.S.C. § 1292(b).

First, plaintiffs fail to show a controlling question of law. While plaintiffs focus on the statutory question this case implicates, that is not enough to show that the certified order addresses a question of law. The order the district court certified simply applied its interpretation of the DMCA to new facts alleged in the operative complaint, with little analysis of the statutory identity question. That legal question had been resolved months earlier in an order that plaintiffs never sought to certify. And regardless, simply pointing to a legal question is not enough—that legal question must also be *controlling*. Plaintiffs cannot make that showing here because

district court litigation on plaintiffs' remaining claim will proceed regardless of appeal.

Second, plaintiffs do not identify any disagreement on that statutory question. No federal appellate court has addressed the relevant DMCA provision. Instead of identifying any conflicting appellate precedent, plaintiffs trawl through various district court DMCA opinions to create the specter of a conflict. But plaintiffs ignore that many of those decisions considered different legal arguments or turned on distinct facts. Those differences make clear that any disagreement among the district courts is largely illusory. And while plaintiffs themselves clearly disagree with the district court's decision, that disagreement has no bearing on this petition.

Third, plaintiffs cannot show that immediate appeal will materially advance the litigation. In fact, it will have the opposite effect by opening the door to duplicative, if not entirely unnecessary, appellate proceedings. Plaintiffs' arguments to the contrary rely on facts ordinary to litigation. But Section 1292(b) is not designed to review the ordinary case.

Finally, plaintiffs' focus on the "importance" of a legal question is irrelevant. Whether an issue is important is not one of the statutory factors under Section 1292(b). Indeed, courts have given importance exactly the opposite weight: the more important an issue, the more reason it should be reviewed in an ordinary-course appeal on a fully developed record. Such an appeal will surely arise as this case and

others continue to work their way through the district courts. There is no reason for plaintiffs to jump to the front of the line.

### **STATEMENT OF THE CASE**

The DMCA provides that no person shall “intentionally remove or alter any copyright management information” or “distribute ... copies of works ... knowing that copyright management information has been removed or altered.” 17 U.S.C. § 1202(b)(1), (3). Plaintiffs are software developers who allege that Copilot—an artificial intelligence-based coding tool released by co-defendant GitHub and which uses OpenAI’s technology—output modified portions of their software code without including CMI. ECF 200 ¶¶ 205-06. Plaintiffs also allege that OpenAI breached a license agreement by taking the same actions. ECF 200 ¶¶ 245-47. The district court ruled on OpenAI’s first motion to dismiss in May 2023, permitting plaintiffs’ DMCA Sections 1202(b)(1) and (b)(3) theories to proceed but declining to address OpenAI’s argument that the DMCA claims failed because plaintiffs did not allege that CMI had been removed from identical copies of plaintiffs’ works. ECF 95 at 18-21; Ex. A (ECF 53) at 9-10.

Plaintiffs’ First Amended Complaint could allege only that Copilot was outputting at most modified portions of plaintiffs’ code, not identical copies of any works. Ex. B (ECF 135) ¶¶ 96, 103-28. OpenAI moved to dismiss the DMCA claims again, arguing that the statute’s identity requirement was unsatisfied, and



also that plaintiffs had failed to plausibly allege that Copilot actually removed CMI because “[i]n order to get Copilot to output modified copies of portions of Plaintiffs’ source code,” plaintiffs had to “provide[] the majority of the work to Copilot (without CMI) as the input.” ECF 110 at 11-12. The district court agreed with OpenAI that Section 1202(b) requires identity and dismissed the DMCA claims without prejudice. ECF 189 at 14-16. Two months later, plaintiffs asked the district court to reconsider its ruling on the identity requirement; the district court declined. Ex. C (ECF 218); Ex. E (ECF 246).

Plaintiffs’ Second Amended Complaint continued to allege simply that Copilot outputs plaintiffs’ code in “[m]odified [f]ormat” and creates “modified copies” in response to detailed input from plaintiffs, which are not an “exact match” or are merely “functionally equivalent” to plaintiffs’ code. ECF 200 ¶¶ 115, 120, 124-25, 133. Plaintiffs also included new allegations describing GitHub’s code-referencing feature and an unrelated academic study about language models. ECF 200 ¶¶ 104, 146, 208-09. The district court again dismissed the DMCA claims—this time, with prejudice. ECF 253 at 4-6. The district court observed that while some of plaintiffs’ arguments focused on whether “identity is not an element of a Section 1202(b) claim,” it would not “revisit” that question “at length,” “[h]aving twice addressed this issue already,” once in the prior dismissal order and once on reconsideration. ECF 253 at 4. Accordingly, it focused on plaintiffs’ new factual

allegations, deeming them insufficient to support the inference that Copilot would distribute identical copies of plaintiffs' works.

Plaintiffs moved to certify this order (the Third MTD Order) for interlocutory appeal. ECF 268. The district court granted the motion. It construed the question to be certified as “whether Sections 1202(b)(1) and (b)(3) of the DMCA impose an identity requirement,” which the court deemed “a purely legal question of statutory interpretation”—a question that the court had expressly declined to revisit in the Third MTD Order. ECF 282 at 2. It reasoned this question was “controlling” because if it were “decided in Plaintiffs’ favor they would be able to proceed with their DMCA claims.” ECF 282 at 2. The district court also determined there was a substantial ground for difference of opinion, citing three district court decisions that declined to dismiss DMCA claims. ECF 282 at 3. Finally, the district court concluded that immediate appeal was likely to “materially advance the ultimate outcome of the litigation” because it would result in a “final decision” on the question plaintiffs presented. ECF 282 at 3.

### **REASONS THE PETITION SHOULD BE DENIED**

This Court has discretion to review an interlocutory order that a district court has certified as (1) involving a controlling question of law (2) as to which there is substantial ground for difference of opinion and (3) where an immediate appeal from the order may materially advance the ultimate termination of the litigation.

28 U.S.C. § 1292(b). “Because the “requirements of § 1292(b) are jurisdictional, if th[e] appeal does not present circumstances satisfying the statutory prerequisites for granting certification, this court cannot allow” it. *Couch v. Telescope Inc.*, 611 F.3d 629, 633 (9th Cir. 2010) (quotation marks omitted). “[E]ven if the district judge certifies the order under § 1292(b), the appellant still has the burden of persuading the court of appeals that exceptional circumstances justify a departure from the basic policy of postponing appellate review until after the entry of a final judgment.” *Coopers & Lybrand v. Livesay*, 437 U.S. 463, 475 (1978) (quotation marks omitted), *superseded on other grounds by rule as stated in Microsoft Corp. v. Baker*, 582 U.S. 23 (2017). This Court has discretion to deny the request “for any reason.” *Id.*

None of Section 1292(b)’s requirements is satisfied here. In concluding otherwise, the district court misconstrued the statutory standards and other district court decisions. And even if those standards were satisfied, there are good reasons for the Court to exercise its discretion to decline review.

The Court should deny the petition.

# **I. ANY QUESTION OF LAW PRESENTED HERE IS NOT CONTROLLING**

As plaintiffs agree (Pet. 6), a “question of law” must be a pure legal question that can be answered without examining the facts of a case. *Steering Comm. v. United States*, 6 F.3d 572, 575-76 (9th Cir. 1993). By contrast, the application of a



legal “standard to the facts of th[e] case ... is a mixed question of law and fact” that is not appropriate for interlocutory review. *Id.* at 575.

Plaintiffs argue that this requirement is satisfied because “whether § 1202(b) requires CMI be removed or altered from identical copies” of a work is a “pure question[] of law.” Pet. 6. But the Third MTD Order that the district court certified answered a different question: whether the new allegations in plaintiffs’ Second Amended Complaint adequately pleaded a violation of the DMCA. That question was a “mixed” one, requiring the court to examine specific allegations about a prototype software code-referencing feature and an academic study. ECF 253 at 2-3, 5-6 (concluding that the code-referencing feature “does not make it more likely that Copilot would ever output an identical copy of Plaintiffs’ works” and that “the Carlini Study does nothing to rehabilitate Plaintiffs’ own concession that ... Copilot’s suggestions are a modification”) (emphasis and quotation marks omitted). Indeed, plaintiffs expressly focused on the DMCA’s “application to software code in particular.” Ex. D (ECF 235) at 5. Thus, rather than addressing “a pure or abstract question about the” DMCA itself, the Third MTD Order assessed “whether the specific facts alleged by these particular plaintiffs state ... claims for relief under the” DMCA—a form of analysis that is ill-suited for interlocutory review. *See Mamani v. Berzain*, 825 F.3d 1304, 1313 (11th Cir. 2016) (exercising discretion to

decline review when the relevant question turned on the application of the Torture Victims Protection Act to “224 paragraphs of allegations spanning 55 pages”).

In seeking certification and before this Court, plaintiffs frame the relevant question as whether the DMCA requires identity. But the district court answered that question in a separate order, issued more than six months before the Third MTD order, and which plaintiffs never sought to certify. ECF 253 at 4-6. To the extent plaintiffs seek review of that months-old ruling, their request is “inexcusably dilatory.” *Richardson Elecs., Ltd. v. Panache Broad. of Pa., Inc.*, 202 F.3d 957, 958 (7th Cir. 2000) (Posner, J.) (declining certification based on two-month delay).

The result would be the same even if the Third MTD Order cleanly presented that legal question. Plaintiffs emphasize that their “identity” question is one of pure statutory interpretation and claim this is enough to satisfy this Section 1292(b) prong. Pet. 7. Plaintiffs are wrong. For one thing, other appellate courts have rejected a standard that looks exclusively at whether the certified question is purely legal, *Mamani*, 825 F.3d at 1313, as that would dramatically expand the scope of issues appropriate for interlocutory review.<sup>1</sup> For another, any question of law must

---

<sup>1</sup> Moreover, the decisions plaintiffs cite did not rule on a Section 1292(b) petition, and thus offer no guidance as to when a question of statutory interpretation might satisfy Section 1292(b)’s requirements. *Joffe v. Google, Inc.*, 746 F.3d 920, 923-24 (9th Cir. 2013); *Sec’y, U.S. Dep’t of Lab. v. Preston*, 873 F.3d 877, 880 (11th Cir. 2017); *Fast v. Applebee’s Int’l, Inc.*, 638 F.3d 872, 785-76 (8th Cir. 2011).

also be “controlling”—meaning that reversal on appeal would “avoid protracted and expensive litigation.” *In re Cement Antitrust Litig.*, 673 F.2d 1020, 1026 (9th Cir. 1981). Plaintiffs’ petition is silent on this requirement.

Perhaps for good reason. The district court deemed a question “controlling” where if it were “decided in Plaintiffs’ favor they would be able to proceed.” ECF 282 at 2. But that circular reasoning would embrace a wide swath of Section 1292(b) petitions, as Section 1292(b) petitioners often seek reversal of a ruling on an issue they wish to pursue in the district court.

And plaintiffs themselves have no answer to the “controlling” factor because it (dovetailing with the third statutory factor, *see* 16 Fed. Prac. & Proc. Juris. § 3930 (3d ed.)) looks to whether immediate resolution of the question will avoid unnecessary and costly litigation. *In re Cement Antitrust Litig.*, 673 F.2d at 1026. But plaintiffs’ breach-of-contract claim, which was not dismissed, will have to be litigated regardless of the fate of the DMCA claims. Plaintiffs’ contract claim is, moreover, based on the same facts as the DMCA claims. ECF 200 ¶¶ 245-47. And plaintiffs have taken the position that “[m]any (if not all) of the discovery requests relevant to the DMCA claims and the breach of contract claim are co-extensive” such that the “narrowing of the claims in this case does not alter the relevance of the discovery sought.” Ex. F (ECF 274-2). Thus, the DMCA identity question is not controlling because plaintiffs’ own position is that its appeal will not expedite

the district court litigation. *Matter of Hamilton*, 122 F.3d 13, 15 (7th Cir. 1997); *Rabin v. PricewaterhouseCoopers LLP*, No. 16-cv-02276, 2017 WL 11662124, at \*2 (N.D. Cal. Apr. 17, 2017) (question was “not controlling” when “much of the discovery on” that issue “would overlap with the discovery required for” other claims).

For similar reasons, review would be unwarranted even if this Section 1292(b) requirement were satisfied (which it is not). The presence of potentially overlapping claims makes it all the more sensible for this Court to await review until a final judgment disposes of them all. *Cf. In re Prudential Lines, Inc.*, 59 F.3d 327, 332-33 (2d Cir. 1995) (declining interlocutory review because “it would be more efficient” to hear appeal concurrently with another appeal where both “require construction” of the same “policies”).

## **II. THE PETITION DOES NOT IDENTIFY A SUBSTANTIAL GROUND FOR A DIFFERENCE IN OPINION**

Section 1292(b) is “not intended merely to provide review of difficult rulings in hard cases.” *United States Rubber Co. v. Wright*, 359 F.2d 784, 785 (9th Cir. 1966) (per curiam). Rather, interlocutory review is appropriate only if “the circuits are in dispute on the question and the court of appeals of the circuit has not spoken on the point, if complicated questions arise under foreign law, or if novel and difficult questions of first impression are presented.” *Couch*, 611 F.3d at 633.



Although plaintiffs—like the district court (ECF 282 at 3)—emphasize this factor, it provides no support for interlocutory review.

Neither plaintiffs nor the district court identify any inter-circuit “dispute” on this issue that this Court may need to resolve. Indeed, it appears that no circuit has spoken directly on the question. And “the mere presence of a disputed issue that is a question of first impression, standing alone, is insufficient to demonstrate a substantial ground for difference of opinion.” *Couch*, 611 F.3d at 634 (quoting *In re Flor*, 79 F.3d 281, 284 (2d Cir. 1996)).

Instead, analysis of the question has been confined to the district courts, which have concluded that “no DMCA violation exists where the works are not identical.” *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. 20-cv-1931, 2020 WL 5991503, at \*6 (C.D. Cal. Aug. 14, 2020). For instance, DMCA claims have been dismissed when a defendant’s “[v]irtually identical plans could have been created by redrawing” the plaintiff’s “plans and not including” plaintiff’s “copyright management information,” because that situation “would not involve any removal or alteration of copyright management information from” the plaintiff’s “original work.” *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-cv-0496, 2014 WL 5798282, at \*5 (D. Haw. Nov. 7, 2014), *aff’d*, 700 F. App’x 674, 675 (9th Cir. 2017) (affirming “for the reasons stated in the district court’s order”). Similarly, district courts have rejected DMCA claims when defendants distributed “knockoff



products” with mere “similarities” to the plaintiff’s products, *Dolls Kill, Inc. v. Zoetop Bus. Co.*, No. 22-cv-1463, 2022 WL 16961477, at \*3-4 (C.D. Cal. Aug. 25, 2022); when a defendant modified the “opacity and position” of copyrighted designs, *O’Neal v. Sideshow, Inc.*, 583 F. Supp. 3d 1282, 1287 (C.D. Cal. 2022); when “information from” a plaintiff’s academic “courses was allegedly copied into a different form and then incorporated into the [defendant’s] note packages,” *Faulkner Press, L.L.C. v. Class Notes, L.L.C.*, 756 F. Supp. 2d 1352, 1359 (N.D. Fla. 2010); and when a defendant’s work “drew upon various materials” provided by a plaintiff, *Fischer v. Forrest*, 286 F. Supp. 3d 590, 610 (S.D.N.Y. 2018) (approvingly citing *Frost-Tsuji* and *Faulkner*), *aff’d*, 968 F.3d 216 (2d Cir. 2020).

Plaintiffs and the district court here purported to find a “split” on this question based on a handful of other district court decisions (most outside this Circuit). But those decisions either answered different legal questions or confronted different facts. At the very most, they show that “settled law might be applied differently,” which “does not establish a substantial ground for difference of opinion.” *Couch*, 611 F.3d at 633.

For instance, the district court claimed to have diverged from a Western District of Texas decision, *ADR International Ltd. v. Institute for Supply Management Inc.* ECF 282 at 3; *see* 667 F. Supp. 3d 411, 428 (W.D. Tex. 2023) (allowing DMCA claims to proceed based on allegations that defendant “repeatedly

copied the language and presentation from ADR’s slides while making superficial changes”). But *ADR* considered the relevance of another Western District of Texas decision, *Kipp Flores Architects, LLC v. Pradera SFR, LLC*. See *ADR*, 667 F. Supp. 3d at 428-29; *Kipp Flores*, No. 21-cv-00673, 2022 WL 1105751, at \*3 (W.D. Tex. Apr. 13, 2022). *Kipp Flores* had dismissed DMCA claims on the ground that “[r]emoval of CMI from a copyrighted work is not the same as the failure to add CMI to a nonidentical rendition or a derivative of the protected work.” *ADR*, 667 F. Supp. 3d at 428 (quoting *Kipp Flores*, 2022 WL 1105751, at \*3). The *ADR* court distinguished *Kipp Flores* because *ADR* involved different facts: the plaintiff there pleaded not that the defendant created a new work from the plaintiff’s work, but instead that the defendant “reproduced [p]laintiff’s training materials”—specifically, PowerPoint slides—“without its CMI and replaced it with” defendant’s CMI. *ADR*, 667 F. Supp. at 428-29 (emphasis added). Thus, contrary to the district court’s understanding, *ADR* would not support DMCA claims here, where Copilot’s output is alleged to be a “modification based on a copy of” plaintiffs’ code, not a verbatim reproduction of that code. ECF 200 ¶ 120 (emphasis added).

The other decisions that plaintiffs cite—many of which are unpublished—do not show disagreement, either. In most, the district court did not rule on the identity requirement at all. For instance, in *Software Pricing Partners, LLC v. Geisman*, there was no dispute about identity, as the *only* way in which the

defendant had “altered” plaintiff’s works was by removing or falsifying CMI. No. 19-cv-0195, 2022 WL 3971292, at \*5, \*10 (W.D.N.C. Aug. 31, 2022) (observing that defendant “sent an electronic copy of the same [copyrighted] document ... [but] removed the indications that [plaintiff] owned the document and inserted his own name”). The district court did not need to consider (let alone hold) whether the work from which CMI was removed need have only a “substantial similarity” to the original. *Contra* Pet. 9. Likewise, the question presented in both *Bounce Exchange, Inc. v. Zeus Enterprises, Ltd.* and *Splunk, Inc. v. Cribl, Inc.* was whether the plaintiffs had pleaded the *existence* of CMI—not whether the works defendants distributed absent that CMI were *identical* copies of the plaintiffs’ works. *See Bounce*, No. 15-cv-3268, 2015 WL 8579023, at \*2-4 (S.D.N.Y. Dec. 9, 2015) (ruling that certain “terms Bounce incorporated into its source code ... satisfy the statutory definition of CMI”); *Splunk*, 662 F. Supp. 3d 1029, 1053 (N.D. Cal. 2023) (rejecting argument that plaintiff “failed to plead copyright management information ... with the specificity that courts in this district have required”). Similarly, in *Real World Media LLC v. Daily Caller, Inc.*, the court’s ruling did not turn on any “identical[ity]” issues, but instead on the “narrower, distinct question” of whether “an *exact copy* of *portions* of an original work” can support a DMCA claim. No. 23-cv-1654, 2024 WL 3835351, at \*10 (D.D.C. Aug. 14, 2024). But that is not the question for which plaintiffs seek certification. Pet. 2.

Other decisions, when analyzed more fully, reveal that any disagreement is overblown. For instance, *Oracle International Corp. v. Rimini Street, Inc.*’s statement that DMCA liability can lie for modification of code “substantially similar” to a plaintiff’s code quoted another decision, *Enterprise Technology Holdings, Inc. v. Noveon Systems, Inc.* See *Oracle*, No. 14-cv-1699, 2023 WL 4706127, at \*82 (D. Nev. July 24, 2023). In *Enterprise*, the modifications that supported DMCA liability were changes to “name” and “author” information in source code—that is, *changes to CMI*. No. 05-cv-2236, 2008 WL 11338356, at \*16 (S.D. Cal. July 29, 2008). In other words, just like in *Software Pricing Partners* (*supra* pp. 13-14), the only distinction between a plaintiff’s original work and the work distributed by defendants was the removal or alteration of CMI. That standard is *consistent with*, not different from, the standard the district court applied in this case. ECF 189 at 14-16; ECF 253 at 4-6.

Still other decisions merely claim to observe the existence of a split, but without demonstrating any real disagreement. Pet. 10-11 (discussing *Beijing Meishe Network Tech. Co. v. TikTok Inc.*, No. 23-cv-06012, 2024 WL 3522196, at \*9 (N.D. Cal. July 23, 2024), and *Andersen v. Stability AI Ltd.*, No. 23-cv-0201, 2024 WL 3823234, at \*8 (N.D. Cal. Aug. 12, 2024)). Those untethered observations are not sufficient to support interlocutory review.



Unable to satisfy this statutory requirement, plaintiffs resort to attacking the merits of the Third MTD Order. Pet. 16-18. But “[a] party’s strong disagreement with the Court’s ruling is not sufficient for there to be a substantial ground for difference” either. *Couch*, 611 F.3d at 633 (alteration in original) (quotation marks omitted).

And plaintiffs’ statutory reading is wrong regardless. Whether the DMCA requires identity does not turn on the presence or absence of the word “identical” in Section 1202(b); rather, the identity concept is conveyed in the language that CMI be “removed” from or “altered” in a work. 17 U.S.C. § 1202(b). CMI cannot be removed from or altered in a work that is newly created based on a plaintiff’s original (such as a derivative work) because that new work never contained CMI to begin with. *See, e.g., Frost-Tsuji*, 2014 WL 5798282, at \*7 (no genuine issue of fact as to whether defendant “*removed* Frost-Tsuji’s copyright management information” based on allegations that defendant “created shop drawings based on Frost-Tsuji’s work and left out Frost-Tsuji’s copyright management information in the process”) (emphasis by court). This concept can, but need not, be conveyed using the word “identical.”

### **III. IMMEDIATE APPEAL WOULD NOT MATERIALLY ADVANCE THE LITIGATION**

The “materially advance prong is satisfied when the resolution of the question may appreciably shorten the time, effort, or expense of conducting the district court



proceedings.” *ICTSI Or., Inc. v. Int’l Longshore & Warehouse Union*, 22 F.4th 1125, 1131 (9th Cir. 2022) (quotation marks omitted). For the same reasons Plaintiffs fail to show a “controlling” question, they also fail to show the prospect of material advancement. *Supra* pp. 9-10. Neither the district court’s analysis, nor plaintiffs’ new arguments, shows otherwise, as all focus on aspects of litigation that will be true in every case.

To begin: The district court reasoned that the “materially advance” requirement was satisfied because a ruling from this Court would result in a “final decision” on the question plaintiffs present. ECF 282 at 3. That conflates resolution of a claim—which will always happen on an appeal, interlocutory or not—with the efficiency and speed concerns that animate this statutory requirement. As other appellate decisions recognize, “[n]o doubt, an immediate appeal has the potential to alter th[e] litigation’s path and outcome, just as an appeal of almost any controlling question of law would.” *In re Somberg*, 31 F.4th 1006, 1009 (6th Cir. 2022). But the question is whether “immediate review” will “accelerate its end *in this instance*.” *Id.* (emphasis added). The district court did not answer this latter, critical question.<sup>2</sup>

---

<sup>2</sup> Plaintiffs characterize the district court’s order as concluding that “resolving this legal question at an early stage would prevent the parties from incurring significant litigation costs on issues that might ultimately be rendered moot by a later ruling from the Ninth Circuit.” Pet. 5. The order mentioned nothing about mootness. ECF 282 at 3-4. Nor did the district court say anything about “discovery, expert testimony, and class certification proceedings.” Pet. 5.

In fact, interlocutory appeal here will only *undermine* efficiency interests. If this Court “den[ies] permission to appeal now,” the district court can address the contract claim, and this Court may “review” plaintiffs’ “claims all at once at the end of the proceedings below”—the normal way dismissed claims are reviewed. *Id.* By contrast, if this Court “grant[s] permission” to appeal, then “regardless of what” it concludes about the DMCA claims, “the district court would need to address” the contract claim, and this “case could easily come back to this court after that.” *Id.* The strongly held “interest against piecemeal appeals” counsels against such a bifurcated procedure. *Id.* (declining interlocutory review); *cf. Wood v. GCC Bend, LLC*, 422 F.3d 873, 879-80 (9th Cir. 2005) (deeming Rule 54(b) certification inappropriate when it would result in “piecemeal appeals with respect to the same set of facts”).

Plaintiffs’ arguments to this Court to the contrary are unpersuasive. Most of the circumstances they claim support their “materially advance” argument are common to all or most cases: the presence of damages; the early procedural posture; and the final word that appellate review always provides. Those circumstances describe ordinary litigation—not the rare case of an appeal suitable for interlocutory review.

For instance, plaintiffs contend that their DMCA claims—the purported “heart” of their case—promise a large amount of damages. Plaintiffs mistakenly

rely on *Sterk v. Redbox Automated Retail, LLC*, 672 F.3d 535, 536 (7th Cir. 2012). There, interlocutory appeal would serve efficiency gains because the “main” claim had been *preserved* at the district court, such that “completion of the litigation will take longer” absent an interlocutory appeal that might eliminate the need for that litigation. The opposite is true in this case, where appeal could only expand the existing litigation. *Supra* pp. 9-10, 17-18. Regardless, whatever the damages for plaintiffs’ DMCA claims might be (and plaintiffs never point this Court to any allegations establishing “tens of thousands” of “individual violations,” Pet. 13), the idea that high potential damages can justify interlocutory review would contravene the principle that such review is reserved for “exceptional” cases. *In re Cement Antitrust Litig.*, 673 F.2d at 1027. Rather, as plaintiffs’ cited cases explain, the role a legal issue plays within a larger case is relevant to this statutory requirement when reversal on that issue could eliminate the case entirely or fundamentally alter its character. *U.S. ex rel. Huangyan Import & Export Corp. v. Nature’s Farm Prods., Inc.*, 370 F. Supp. 2d 993, 1005 (N.D. Cal. 2005) (certifying order because, depending on resolution of issues, plaintiffs might have “no claim at all or might be in the wrong court altogether”) (cited at Pet. 13). As shown above, that is not true here. *Supra* pp. 9-10, 17-18.

Plaintiffs also hypothesize that the certainty a decision from this Court will provide would encourage settlement. Pet. 14. That is, again, true for every appeal.

Plaintiffs offer no reason to think that either defendant's willingness to settle *this* case turns on a favorable ruling on their DMCA claims. Plaintiffs' reliance on the supposedly greater settlement prospects in class actions (at 14) is also premature, given the absence of any motions or rulings on class certification.

Finally, plaintiffs argue that interlocutory appeal will benefit litigants in "other cases challenging AI models ... asserting similar DMCA claims." Pet. 15. But plaintiffs apparently ask the Court to take their word for it, citing no cases that would be so impacted. And even if they had, plaintiffs point to no decision from this Court indicating that the purported need for such generalized guidance in other cases is an element of the "materially advance" prong.

At bottom, plaintiffs' request for interlocutory review amounts to little more than a high-pitched claim that the issue bears on "the future of digital copyright enforcement." Pet. 16. But that is precisely why this Court should wait for an ordinary-course appeal from a final judgment, with a fully developed record. As the foregoing discussion shows, it is far from clear at this stage of the case whether and how the viability of plaintiffs' claims turns on the application of the DMCA to the unique facts of this case. *Supra* p. 7 (describing district court's consideration of allegations new to the Second Amended Complaint). And there is, at best, no good reason to believe that the statutory-interpretation issue that drove the dismissal of plaintiffs' claims is the same as those raised in other district court decisions or a



distinct question. *Supra* pp. 12-14 (describing district courts’ differing formulations of the DMCA issues). Indeed, it might well be that the identity question is not determinative *in this very case* because plaintiffs’ DMCA claims fail for other reasons. *Supra* pp. 3-4 (describing OpenAI’s alternative argument for why DMCA claims failed, which the district court has not yet addressed).

Courts have deemed interlocutory appeal premature under similar circumstances, where the relevance and scope of legal questions turned on such unresolved issues. *E.g.*, *Slade v. Shearson, Hammill & Co.*, 517 F.2d 398, 400 (2d Cir. 1974) (certification inappropriate when “factual questions” would “have a bearing on what is the precise question of law presented by the case,” the resolution of which “may make the question ‘certified’ not controlling in any event”); *S.B.L. ex rel. T.B. v. Evans*, 80 F.3d 307, 311 (8th Cir. 1996) (declining to review on interlocutory appeal the question of institutional liability under Title IX when the scope of that question turned on unresolved credibility issues); *Garcia v. Johanns*, 444 F.3d 625, 637 (D.C. Cir. 2006) (declining to certify when “claim will benefit from further development in the district court”). This reflects the reality that on “an appeal from a final decision, the record and judgment from the trial court” will “have narrowed the legal questions presented and provide[]” this Court “with more complete factual findings.” *Int’l Soc’y for Krishna Consciousness, Inc. v. Air Canada*, 727 F.2d 253, 257 (2d Cir. 1984). Even if plaintiffs are correct about the



“importance and implications of” this DMCA issue, that is a reason this Court should wait to address it. *Id.*; accord *Koehler v. Bank of Bermuda Ltd.*, 101 F.3d 863, 867 (2d Cir. 1996) (Section 1292(b) certification “was improvidently granted” when “premature treatment of” a personal jurisdiction question would “risk the development of unsound precedent in an area with great practical implications for the banking industry”).

### CONCLUSION

The petition should be denied.

Dated: October 17, 2024

Respectfully submitted,

ALEXANDRA M. AVVOCATO  
MORRISON & FOERSTER LLP  
250 West 55th Street  
New York, NY 10019

s/ Aileen M. McGrath

---

AILEEN M. MCGRATH  
JOSEPH C. GRATZ  
TIFFANY CHEUNG  
MORRISON & FOERSTER LLP  
425 Market Street  
San Francisco, CA 94105  
Tel.: (415) 268-6153  
AMcGrath@mofo.com

*Counsel for OpenAI Inc., OpenAI, L.P., OpenAI OpCo, L.L.C., OpenAI GP, L.L.C., OpenAI Startup Fund GP I, L.L.C., OpenAI Startup Fund I, L.P., OpenAI Startup Fund Management, LLC, OpenAI LLC, OpenAI Global, LLC, OAI Corporation, OpenAI Holdings, LLC, OpenAI HoldCo LLC, OpenAI Startup Fund SPV I, LP, and OpenAI Startup Fund SPV GP I LLC*

### **CERTIFICATE OF COMPLIANCE**

I hereby certify that the foregoing response complies with the type-volume limitations of Ninth Circuit Rules 5-2(b) and 32-3 because its total number of words—5,163—divided by 280 does not exceed 20 pages.

This response complies with the typeface requirement of Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. P. 32(a)(6) because it has been prepared in a proportionally spaced typeface, including serifs, using Microsoft Word 2016 in Times New Roman 14-point font.

Dated: October 17, 2024

s/ Aileen M. McGrath  
\_\_\_\_\_  
Aileen M. McGrath

### **CERTIFICATE OF SERVICE**

I hereby certify that I electronically filed the foregoing with the Clerk of the Court for the United States Court of Appeals for the Ninth Circuit by using the ACMS system on October 17, 2024.

I certify that all participants in the case are registered ACMS users and that service will be accomplished by the ACMS system.

Dated: October 17, 2024

\_\_\_\_\_  
s/ Aileen M. McGrath  
Aileen M. McGrath

# EXHIBIT A

MICHAEL A. JACOBS (SBN 111664)  
MJacobs@mofo.com  
JOSEPH C. GRATZ (SBN 240676)  
JGratz@mofo.com  
TIFFANY CHEUNG (SBN 211497)  
TCheung@mofo.com  
MORRISON & FOERSTER LLP  
425 Market Street  
San Francisco, California 94105-2482  
Telephone: (415) 268-7000  
Facsimile: (415) 268-7522  
[CAPTION PAGE CONTINUED ON NEXT PAGE]

Attorneys for Defendants OPENAI, INC., a Delaware nonprofit corporation, OPENAI, L.P., a Delaware limited partnership, OPENAI GP, L.L.C., a Delaware limited liability company, OPENAI STARTUP FUND GP I, L.L.C., a Delaware limited liability company, OPENAI STARTUP FUND I, L.P., a Delaware limited partnership, OPENAI STARTUP FUND MANAGEMENT, LLC, a Delaware limited liability company

**UNITED STATES DISTRICT COURT**  
**NORTHERN DISTRICT OF CALIFORNIA**  
**SAN FRANCISCO DIVISION**

J. DOE 1 and J. DOE 2, individually and on  
behalf of all others similarly situated,

Plaintiffs,

v.

GITHUB, INC., a Delaware corporation;  
MICROSOFT CORPORATION, a Washington  
corporation; OPENAI, INC., a Delaware  
nonprofit corporation; OPENAI, L.P., a Delaware  
limited partnership; OPENAI GP, L.L.C., a  
Delaware limited liability company; OPENAI  
STARTUP FUND GP I, L.L.C., a Delaware  
limited liability company; OPENAI STARTUP  
FUND I, L.P., a Delaware limited partnership;  
OPENAI STARTUP FUND MANAGEMENT,  
LLC, a Delaware limited liability company,

Defendants.

Case No. 4:22-cv-06823-JST  
4:22-cv-07074-JST

Hon. Jon S. Tigar

**CLASS ACTION**

**DEFENDANTS OPENAI, INC.,  
OPENAI, L.P., OPENAI GP, L.L.C.,  
OPENAI STARTUP FUND GP I,  
L.L.C., OPENAI STARTUP FUND I,  
L.P. AND OPENAI STARTUP FUND  
MANAGEMENT, LLC'S NOTICE  
OF MOTION AND MOTION TO  
DISMISS COMPLAINT;  
MEMORANDUM OF POINTS AND  
AUTHORITIES**

Date: May 4, 2023  
Time: 2:00 p.m.  
Courtroom: 6



ROSE S. LEE (SBN 294658)  
RoseLee@mofo.com  
MORRISON & FOERSTER LLP  
707 Wilshire Boulevard  
Los Angeles, California 90017-3543  
Telephone: (213) 892-5200  
Facsimile: (213) 892-5454

**NOTICE OF MOTION AND MOTION**

TO ALL PARTIES AND THEIR ATTORNEYS OF RECORD:

PLEASE TAKE NOTICE that on May 4, 2023 at 2:00 p.m., or at a different time and date set by the Court, Defendants OPENAI, INC., OPENAI, L.P., OPENAI GP, L.L.C., OPENAI STARTUP FUND GP I, L.L.C., OPENAI STARTUP FUND I, L.P. AND OPENAI STARTUP FUND MANAGEMENT, LLC (hereinafter “OPENAI ENTITIES”), by and through counsel, will and hereby do move the Court to dismiss all claims asserted in the Complaint against the OpenAI Entities: (1) violation of the Digital Millennium Copyright Act (“DMCA”), 17 U.S.C. § 1202, *et seq.*; (2) breach of contract; (3) “tortious interference in a contractual relationship”; (4) false designation of origin under the Lanham Act, 15 U.S.C. § 1125; (5) unjust enrichment under common law and California Business & Professions Code § 17200, *et seq.*; (6) unfair competition under common law, the Lanham Act, 15 U.S.C. § 1125, and California Business & Professions Code § 17200, *et seq.*; (7) violations of the California Consumer Privacy Act (“CCPA”), California Civil Code § 1798.150; (8) negligence; (9) civil conspiracy; and (10) declaratory relief.

This Motion is made pursuant to Federal Rules of Civil Procedure 12(b)(1) and 12(b)(6), and is based upon this Notice of Motion and Motion, the Memorandum of Points and Authorities included herewith, the Declaration of Michael Jacobs and attached exhibits, the Proposed Order submitted herewith, all pleadings and papers on file in this action, and such further evidence that may be submitted to the Court or before the hearing.

## TABLE OF CONTENTS

		Page
1		
2		
3	I. INTRODUCTION .....	1
4	II. PLAINTIFFS’ ALLEGATIONS .....	2
5	III. LEGAL STANDARD .....	3
6	A. Motion to Dismiss for Lack of Subject Matter Jurisdiction.....	3
7	B. Motion to Dismiss for Failure to State a Claim Under Rule 12(b)(6). ....	4
8	IV. ARGUMENT .....	5
9	A. The Complaint Fails for Reasons Applicable to All Causes of Action.....	5
10	1. Plaintiffs Lack Article III Standing to Assert Their Claims. ....	5
11	2. Plaintiffs Have Failed to Obtain Leave to Proceed Anonymously. ....	6
12	3. The Complaint’s Undifferentiated Allegations Against the Six OpenAI Entities Fail to Satisfy Pleading Requirements. ....	7
13	B. The Copyright Act Preempts Several State Law Causes of Action. ....	8
14	C. Plaintiffs’ Claims Fail for Reasons Specific to Each Claim. ....	9
15	1. Plaintiffs’ DMCA Claim Should be Dismissed. ....	9
16	a. Plaintiffs Have Not Properly Pled a Claim for Removal of CMI. ....	9
17	b. Plaintiffs Have Failed to Plead a Claim for Distributing Copies of Works from Which CMI Has Been Removed. ....	12
18	c. Plaintiffs Have Failed to Show that OpenAI Has Conveyed Any False CMI in Connection with Copilot Outputs. ....	13
19	2. Plaintiffs’ Breach of Contract Claim Fails.....	13
20	a. Plaintiffs Have Not Sufficiently Pled Existence of a Contract. ....	14
21	b. Plaintiffs Fail to Allege Facts Demonstrating the Contractual Provisions OpenAI Entities Allegedly Breached.....	14
22	3. The Claim for Tortious Interference in Contractual Relationship Fails. ....	15
23	4. Plaintiffs Fail to Allege a False Designation of Origin Claim. ....	16
24	5. Plaintiffs Fail to State a Claim for Unjust Enrichment. ....	17
25	6. Plaintiffs Fail to State an Unfair Competition Claim.....	18
26	7. Plaintiffs Fail to Adequately Plead A Violation of the CCPA.....	19
27	8. Plaintiffs Fail to State a Claim for Negligence. ....	21
28	9. Plaintiffs Fail to State a Civil Conspiracy Claim. ....	24
	10. Plaintiffs Fail to State a Claim for Declaratory Relief.....	25
	V. CONCLUSION .....	25

**TABLE OF AUTHORITIES**

**Page(s)**

**Cases**

<i>AccuImage Diagnostics Corp. v. Terarecon, Inc.</i> , 260 F. Supp. 2d 941 (N.D. Cal. 2003) .....	24, 25
<i>Agence France Presse v. Morel</i> , 769 F. Supp. 2d 295 (S.D.N.Y. 2011).....	16
<i>Aguilar v. Hartford Accident &amp; Indem. Co.</i> , No. CV 18-8123-R, 2019 WL 2912861 (C.D. Cal. Mar. 13, 2019) .....	21, 23
<i>Alsheikh v. Lew</i> , No. 3:15-CV-03601-JST, 2016 WL 1394338 (N.D. Cal. Apr. 7, 2016) .....	6
<i>Anderson v. Kimpton Hotel &amp; Rest. Grp., LLC</i> , No. 19-CV-01860-MMC, 2019 WL 3753308 (N.D. Cal. Aug. 8, 2019) .....	20, 22
<i>Applied Equip. Corp. v. Litton Saudi Arabia Ltd.</i> , 7 Cal. 4th 503 (1994) .....	24
<i>Ashcroft v. Iqbal</i> , 556 U.S. 662 (2009) .....	4
<i>Astiana v. Hain Celestial Grp., Inc.</i> , 783 F.3d 753, 762 (9th Cir. 2015).....	17
<i>Authors Guild v. Google, Inc.</i> , 804 F.3d 202 (2d Cir. 2015).....	12
<i>Bell Atl. Corp. v. Twombly</i> , 550 U.S. 544 (2007) .....	4
<i>Birdsong v. Apple, Inc.</i> , 590 F.3d 955 (9th Cir. 2009).....	5
<i>Cetacean Cmty. v. Bush</i> , 386 F.3d 1169 (9th Cir. 2004).....	4
<i>Chandler v. State Farm Mut. Auto Ins. Co.</i> , 598 F.3d 1115 (9th Cir. 2010).....	4
<i>Corona v. Sony Pictures Ent., Inc.</i> , No. 14-CV-09600 RGK (Ex), 2015 WL 3916744 (C.D. Cal. June 15, 2015).....	23
<i>CSI Elec. Contractors, Inc. v. Zimmer Am. Corp.</i> , No. CV 12-10876-CAS (AJWx), 2013 WL 1249021 (C.D. Cal. Mar. 25, 2013) .....	14



1	<i>Dastar Corp. v. Twentieth Century Fox Film Corp.</i> ,	
2	539 U.S. 23 (2003).....	16
3	<i>Davis v. FEC</i> ,	
4	554 U.S. 724 (2008).....	5
5	<i>Del Madera Props. v. Rhodes &amp; Gardner, Inc.</i> ,	
6	820 F.2d 973 (9th Cir. 1987).....	8
7	<i>Doe v. Kamehameha Schs./Bernice Pauahi Bishop Est.</i> ,	
8	596 F.3d 1036 (9th Cir. 2010).....	6
9	<i>Doe v. UNUM Life Ins. Co. of Am.</i> ,	
10	164 F. Supp. 3d 1140 (N.D. Cal. 2016).....	6
11	<i>Does 1 Thru XXIII v. Advanced Textile Corp.</i> ,	
12	214 F.3d 1058 (9th Cir. 2000).....	6
13	<i>Dolls Kill, Inc. v. Zoetop Business Co.</i> ,	
14	No. 2:22-cv-01463-RGK-MAA, 2022 WL 16961477 (C.D. Cal. Aug. 25, 2022).....	12
15	<i>Eidmann v. Walgreen Co.</i> ,	
16	522 F. Supp. 3d 634 (N.D. Cal. 2021).....	18
17	<i>Firoozye v. Earthlink Network</i> ,	
18	153 F. Supp. 2d 1115 (N.D. Cal. 2001).....	8
19	<i>Free Speech Sys., LLC v. Menzel</i> ,	
20	390 F. Supp. 3d 1162 (N.D. Cal. 2019).....	10
21	<i>Frost-Tsuji Architects v. Highway Inn, Inc.</i> ,	
22	No. CIV. 13-00496 SOM, 2015 WL 263556 (D. Haw. Jan. 21, 2015), <i>aff'd</i> , 700 F. App'x 674 (9th Cir. 2017).....	10
23	<i>Gardiner v. Walmart Inc.</i> ,	
24	No. 20-CV-04618-JSW, 2021 WL 2520103 (N.D. Cal. Mar. 5, 2021).....	23
25	<i>Gen. Am. Life Ins. Co. v. Rana</i> ,	
26	769 F. Supp. 1121 (N.D. Cal. 1991).....	25
27	<i>In re Gilead Scis. Sec. Litig.</i> ,	
28	536 F.3d 1049 (9th Cir. 2008).....	4
	<i>Google LLC v. Oracle Am., Inc.</i> ,	
	141 S. Ct. 1183 (2021).....	12
	<i>Hayden v. Retail Equation, Inc.</i> ,	
	No. SACV2001203DOCDFM, 2022 WL 2254461 (C.D. Cal. May 4, 2022).....	19

1	<i>Ho Myung Moolsan Co. v. Manitou Mineral Water, Inc.,</i>	
2	665 F. Supp. 2d 239 (S.D.N.Y. 2009).....	24
3	<i>Huynh v. Quora, Inc.,</i>	
4	No. 18-CV-07597-BLF, 2020 WL 7408230 (N.D. Cal. June 1, 2020) .....	23
5	<i>Image Online Design, Inc. v. Internet Corp. for Assigned Names &amp; Numbers,</i>	
6	No. CV 12-08968 DDP (JCx), 2013 WL 489899 (C.D. Cal. Feb. 7, 2013) .....	15
7	<i>In re iPhone App. Litig.,</i>	
8	No. 11-MD-022590-LHK, 2011 WL 4403963 (N.D. Cal. Sept. 20, 2011).....	4
9	<i>Johnson v. Weinberger,</i>	
10	851 F.2d 233 (9th Cir. 1988).....	5
11	<i>Kelly v. Arriba Soft Corp.,</i>	
12	77 F. Supp. 2d 1116 (C.D. Cal. 1999),	
13	<i>aff'd and rev'd in part on other grounds</i> , 336 F.3d 811 (9th Cir. 2003) .....	9
14	<i>Kirk Kara Corp. v. W. Stone &amp; Metal Corp.,</i>	
15	No.CV 20-1931-DMG (EX), 2020 WL 5991503 (C.D. Cal. Aug. 14, 2020) .....	10, 12
16	<i>Klein v. Chevron U.S.A., Inc.,</i>	
17	202 Cal. App. 4th 1342 (2012).....	17
18	<i>Kokkonen v. Guardian life Ins. Co. of Am.,</i>	
19	511 U.S. 375 (1994).....	3
20	<i>Korea Supply Co. v. Lockheed Martin Corp.,</i>	
21	29 Cal. 4th 1134 (2003) .....	16
22	<i>Lance v. Coffman,</i>	
23	549 U.S. 437 (2007).....	4
24	<i>Laxalt v. McClatchy,</i>	
25	622 F. Supp. 737 (D. Nev. 1985).....	25
26	<i>Lewis v. Casey,</i>	
27	518 U.S. 343 (1996).....	5
28	<i>Lierboe v. State Farm Mut. Auto. Ins. Co.,</i>	
	350 F.3d 1018 (9th Cir. 2003).....	5
	<i>Lil' Man In the Boat, Inc. v. City and Cnty. of San Francisco,</i>	
	No. C17-CV-00904-JST, 2018 WL 4207260 (N.D. Cal. Sept. 4, 2018) .....	20
	<i>Logan v. Meta Platforms, Inc.,</i>	
	No. 22-cv-1847, 2022 WL 14813836 (N.D. Cal. Oct. 25, 2022) .....	13

1	<i>Low v. LinkedIn Corp.</i> ,	
2	900 F. Supp. 2d 1010 (N.D. Cal. 2012) .....	13
3	<i>Lujan v. Defenders of Wildlife</i> ,	
4	504 U.S. 555 (1992) .....	5
5	<i>Lynwood Invs. CY Ltd. v. Konovalov</i> ,	
6	No. 20-CV-03778-MMC, 2022 WL 3370795 (N.D. Cal. Aug. 16, 2022) .....	12
7	<i>Maag v. U.S. Bank Nat’l Ass’n</i> ,	
8	No. 21-cv-00031-H-LL, 2021 WL 5605278 (S.D. Cal. Apr. 8, 2021) .....	20
9	<i>Malasky v. Esposito</i> ,	
10	No. 16-CV-04102-DMR, 2019 WL 79032, at *10 (N.D. Cal. Jan. 2, 2019), aff’d, 781 F. App’x 643 (9th Cir. 2019) .....	25
11	<i>Maloney v. T3Media, Inc.</i> ,	
12	853 F.3d 1004 (9th Cir. 2017) .....	8
13	<i>Mango v. BuzzFeed, Inc.</i> ,	
14	356 F. Supp. 3d 368 (S.D.N.Y 2019), aff’d, 970 F.3d 167 (2d Cir. 2020) .....	12
15	<i>Mango v. BuzzFeed, Inc.</i> ,	
16	970 F.3d 167 (2d Cir. 2020) .....	11
17	<i>Mayen v. Bank of Am. N.A.</i> ,	
18	No. 14-CV-03757-JST, 2015 WL 179541 (N.D. Cal. Jan. 14, 2015) .....	25
19	<i>Media.net Advert. FZ-LLC v. NetSeer, Inc.</i> ,	
20	156 F. Supp. 3d 1052 (N.D. Cal. 2016) .....	8
21	<i>Mendiondo v. Centinela Hosp. Med. Ctr.</i> ,	
22	521 F.3d 1097 (9th Cir. 2008) .....	4
23	<i>name.space, Inc. v. Internet Corp. for Assigned Names &amp; Numbers</i> ,	
24	No. CV 12-8676 (PLAx), 2013 WL 2151478 (C.D. Cal. Mar. 4, 2013) .....	16
25	<i>O’Neal v. Sideshow, Inc.</i> ,	
26	583 F. Supp. 3d 1282 (C.D. Cal. 2022) .....	9
27	<i>Oppenheimer v. Allvoices, Inc.</i> ,	
28	No. C 14-00499 LB, 2014 WL 2604033 (N.D. Cal. June 10, 2014) .....	17
	<i>Oracle Am., Inc. v. Google Inc.</i> ,	
	872 F. Supp. 2d 974 (N.D. Cal. 2012) .....	12
	<i>Piping Rock Partners, Inc. v. David Lerner Assocs., Inc.</i> ,	
	946 F. Supp. 2d 957 (N.D. Cal. 2013) .....	15



1	<i>Ramirez v. GMAC Mortg.,</i>	
2	No. CV 09-8189 PSG (FFMx), 2010 WL 148167 (C.D. Cal. Jan. 12, 2010).....	14
3	<i>Razuki v. Caliber Home Loans, Inc.,</i>	
4	No. 17CV1718-LAB (WVG), 2018 WL 6018361 (S.D. Cal. Nov. 15, 2018) .....	20
5	<i>Roe v. San Jose Unified Sch. Dist. Bd.,</i>	
6	No. 20-CV-02798-LHK, 2021 WL 292035 (N.D. Cal. Jan. 28, 2021) .....	4
7	<i>Rosas v. City of Santa Rosa,</i>	
8	No. 21-CV-06179-JST, 2022 WL 2158968 (N.D. Cal. June 15, 2022) .....	7
9	<i>Rubio v. U.S. Bank N.A.,</i>	
10	No. C 13-05752 LB, 2014 WL 1318631 (N.D. Cal. Apr. 1, 2014) .....	15
11	<i>Schmitt v. SN Servicing Corp.,</i>	
12	No. 21-CV-03355-WHO, 2021 WL 3493754 (N.D. Cal. Aug. 9, 2021).....	22, 23
13	<i>Sebastian Brown Prods. LLC v. Muzooka Inc.,</i>	
14	143 F. Supp. 3d 1026 (N.D. Cal. 2015) .....	7
15	<i>Sebastian Brown Prods. LLC v. Muzooka Inc.,</i>	
16	No. 15-CV-01720-LHK, 2016 WL 949004 (N.D. Cal. Mar. 14, 2016) .....	19
17	<i>SellPoolSuppliesOnline.com, LLC v. Ugly Pools Ariz., Inc.,</i>	
18	804 F. App'x 668 (9th Cir. 2020) .....	13
19	<i>Silvercrest Realty, Inc. v. Great Am. E&amp;S Ins. Co.,</i>	
20	No. SACV 11-01197-CJC(ANx), 2012 WL 13028094 (C.D. Cal. Apr. 4, 2012).....	18
21	<i>Sonner v. Premier Nutrition Corp.,</i>	
22	971 F.3d 834 (9th Cir. 2020).....	18
23	<i>In re Sony Gaming Networks and Customer Data Sec. Breach Litig.,</i>	
24	996 F. Supp. 2d 942 (S.D. Cal. 2014), <i>order corrected</i> , No. 11MD2258 AJB	
25	(MDD), 2014 WL 12603117 (S.D. Cal. Feb. 10, 2014) .....	21
26	<i>Stevens v. CoreLogic, Inc.,</i>	
27	899 F.3d 666 (9th Cir. 2018).....	9, 11
28	<i>Sulit v. Sound Choice Inc.,</i>	
	No. C06-00045 MJJ, 2006 WL 8442163 (N.D. Cal. Nov. 14, 2006) .....	9
	<i>Sutherland v. Francis,</i>	
	No. 12-CV-05110-LHK, 2014 WL 879697 (N.D. Cal. Mar. 3, 2014) .....	15
	<i>Sybersound Recs., Inc. v. UAV Corp.,</i>	
	517 F.3d 1137 (9th Cir. 2008).....	18



<i>Wasco Prods., Inc. v. Southwall Techs., Inc.</i> , 435 F.3d 989 (9th Cir. 2006).....	24
<i>Wynn v. NBC</i> , 234 F. Supp. 2d 1067 (C.D. Cal. 2002).....	15
<i>Yellowcake, Inc. v. Hyphy Music, Inc.</i> , No. 1:20-CV-0988 AWI BAM, 2021 WL 3052535 (E.D. Cal. July 20, 2021) .....	18
<i>Young v. Facebook, Inc.</i> , 790 F. Supp. 2d 1110 (N.D. Cal. 2011) .....	14
<i>Zepeda v. PayPal, Inc.</i> , 777 F. Supp. 2d 1215 (N.D. Cal. 2011) .....	14
<b>Constitution, Statutes, and Rules</b>	
U.S. Const. art. III .....	3, 4, 5, 6
15 U.S.C. § 1125, <i>et seq.</i> (Lanham Act) .....	<i>passim</i>
17 U.S.C. §§ 1201-1205 (Digital Millennium Copyright Act, (“DMCA”)).....	<i>passim</i>
Cal. Bus & Prof. Code § 17200, <i>et seq.</i> .....	3
Cal. Civ. Code	
§ 1798.82, <i>et seq.</i> .....	22
§ 1798.100, <i>et seq.</i> .....	22
§ 1798.140(i).....	19
§ 1798.150, <i>et. seq.</i> (California Consumer Privacy Act (“CCPA”)).....	<i>passim</i>
§ 1798.155(a) .....	20
Fed. R. Civ. P.	
8.....	5
10(a) .....	6
12(b) .....	4, 5, 6
23(b) .....	3
<b>Other Authorities</b>	
B. Witkin, Summary of California Law, Torts § 44 (9th ed.1988).....	24

1                                    **STATEMENT OF THE ISSUES TO BE DECIDED**

2            This motion raises the following issues:

- 3            1.        **Article III Standing.** Whether the complaint should be dismissed for lack of  
4 standing under Article III of the United States Constitution.
- 5            2.        **Anonymous Pleading.** Whether the complaint should be dismissed because  
6 Plaintiffs have failed to obtain Court approval to proceed anonymously.
- 7            3.        **Minimal Pleading Requirements.** Whether the complaint should be dismissed  
8 for failing to specify the allegations against each OpenAI Entity in violation of pleading  
9 standards.
- 10           4.        **Preemption.** Whether the Copyright Act preempts the state law causes of action  
11 for tortious interference with contract, unjust enrichment, and unfair competition.
- 12           5.        **Digital Millennium Copyright Act, 17 U.S.C. §§ 1201-1205 (Count I).** Whether  
13 this claim should be dismissed for failing to (i) identify specific works from which Copyright  
14 Management Information (“CMI”) was removed, (ii) allege removal of CMI from identical  
15 copies, (iii) allege the requisite intent; (iv) allege distribution of works with removed CMI; or (v)  
16 allege false CMI conveyed in connection with copies of those works.
- 17           6.        **Breach of Contract (Count II).** Whether this claim should be dismissed for  
18 failure to adequately plead that a contract exists and that any OpenAI Entity breached it.
- 19           7.        **Tortious Interference in a Contractual Relationship (Count III).** Whether this  
20 claim should be dismissed for failure to plead that a contract exists between plaintiffs and a third-  
21 party with which any OpenAI Entity allegedly interfered.
- 22           8.        **False Designation (Count V).** Whether this claim should be dismissed because  
23 the Lanham Act does not provide a remedy for false attribution of authorship.
- 24           9.        **Unjust Enrichment (Count VI).** Whether this claim should be dismissed because  
25 there is no standalone cause of action for unjust enrichment and Plaintiffs fail to plead the  
26 required elements.
- 27           10.       **Unfair Competition (Count VII).** Whether this claim should be dismissed for (i)  
28 failure to sufficiently plead a predicate violation or lack of an adequate legal remedy for the

1 Unfair Competition Law (“UCL”) claim under California Business. & Professions Code § 17200  
2 and (ii) an actionable basis for a common law or Lanham Act claim.

3 11. **California Consumer Privacy Act, § 1798.150 (“CCPA”) (Count IX).** Whether  
4 this claim should be dismissed for (i) lack of statutory standing, (ii) failure to provide written  
5 notice of CCPA violations prior to filing, (iii) lack of a private right of action for certain  
6 allegations, and (iv) failure to plead facts showing a violation of OpenAI Entities’ duty to  
7 implement and maintain reasonable security procedures and practices.

8 12. **Negligence (Count X).** Whether this claim should be dismissed for failure to plead  
9 a duty of care owed to Plaintiffs, breach of that duty, causation, and actual damages.

10 13. **Civil Conspiracy (Count XI).** Whether this claim should be dismissed because  
11 civil conspiracy is not an independent cause of action and, in any event, the complaint fails to  
12 adequately plead the role or wrongful acts of each defendant in the alleged conspiracy.

13 14. **Declaratory Relief (Count XII).** Whether this claim should be dismissed because  
14 there is no standalone cause of action for declaratory relief.



**MEMORANDUM OF POINTS AND AUTHORITIES**

**I. INTRODUCTION**

This case is about an emerging field of artificial intelligence known as “Generative AI.” Generative AI systems learn concepts and relationships from large bodies of existing knowledge, and use what they learn to help people create new works. Plaintiffs’ lawsuit challenges two generative AI tools that help people write computer programming code—GitHub Copilot and OpenAI Codex. But Plaintiffs attempt to plead causes of action that don’t actually apply to these tools, rendering the complaint subject to dismissal on multiple grounds.

OpenAI is the creator of Codex and a pioneer in the field of AI. It is an independent company whose mission is to ensure that artificial intelligence, including generative AI systems, benefit all of humanity. OpenAI is governed by a non-profit and its organizational structure limits the economic returns of investors and employees.

Open AI developed Codex using publicly available computer programming code based on theories of how the human brain learns from new information. With that training, Codex generates coding suggestions in response to a person’s requests. A programmer can provide a short text request (*e.g.*, “create a button on a website that lets a user upload a document”), and Codex will generate a coding suggestion to meet the request. Codex’s training allows it to help people write code for common functions and for functions that have never been written before. This capability makes computer programming more accessible to broad segments of the population and makes those who have already learned to program more productive.

The essence of Plaintiffs’ complaint is that rarely—the complaint cites a study reporting 1% of the time—Copilot (and therefore Codex) allegedly generates snippets of code similar to the publicly available code that it learned from, and does so without also generating copyright notices or open source license terms that originally accompanied the code. But Plaintiffs provide no allegation that any code that they authored was used by Codex or generated as a suggestion to a Codex user; they only point to Codex’s abilities to generate common textbook programming functions, such as a function for determining if a number is odd or even. Plaintiffs also do not allege copyright infringement by any OpenAI Entity, but instead allege a grab bag of claims that



1 fail to plead violations of cognizable legal rights.

2 The complaint thus suffers from both threshold defects requiring dismissal of the entire  
3 complaint and specific defects in each pleaded cause of action.

## 4 **II. PLAINTIFFS' ALLEGATIONS<sup>1</sup>**

5 **OpenAI.** While the complaint treats “OpenAI” as a single party, it names six distinct  
6 OpenAI entities as defendants: (1) OpenAI, Inc.; (2) OpenAI, L.P., (3) OpenAI GP, L.L.C.; (4)  
7 OpenAI Startup Fund I, L.P.; (5) OpenAI Startup Fund GP I, L.L.C.; and (6) OpenAI Startup  
8 Fund Management, LLC. The Complaint alleges that OpenAI, Inc. developed Codex, and that  
9 OpenAI, L.P. “co-created” Copilot. (Compl. (Dkt. No. 1) ¶¶ 23-24.) The complaint does not  
10 allege that any of the other OpenAI entities engaged in any of the conduct alleged in the  
11 complaint, but seeks to hold them liable nonetheless. (*Id.* ¶¶ 25-28.)

12 **Plaintiffs.** The named Plaintiffs have filed this complaint under pseudonyms without  
13 leave of Court. Plaintiff J. Doe 3 is a resident of Idaho and Plaintiff J. Doe 4 is a resident of  
14 South Carolina.<sup>2</sup> Plaintiffs allege that they “made available publicly” unspecified “Licensed  
15 Materials” on at least one GitHub repository. (*Id.* ¶¶ 1, 19-20.)

16 Plaintiffs do not allege that their rights associated with code they authored were violated,  
17 nor do they provide a single example of their code they claim to be at issue. The complaint  
18 instead includes excerpts of code functions from third-party textbooks with which Plaintiffs claim  
19 no association: (1) *Eloquent Javascript* by Marijn Haverbeke (*id.* ¶¶ 56-60); (2) *Mastering JS* by  
20 Valeri Karpov (*id.* ¶¶ 71-73); (3) *Think JavaScript* by Matthew X. Curinga et al. (*id.* ¶¶ 74-76).

21 **Plaintiffs' Allegations.** The core of the complaint is that Defendants<sup>3</sup> developed and

---

22  
23 <sup>1</sup> This section is based on the allegations of the complaint, which must be taken as true for  
24 purposes of this motion. By discussing them in this motion, the OpenAI Entities do not admit the  
truth of those allegations.

25 <sup>2</sup> Pursuant to the parties' stipulation and the Court's order, “[t]he Complaint in the Doe 3 Action  
26 shall be deemed the operative Complaint.” (Dkt. No. 46 at 2; Dkt. No. 47.) Accordingly,  
OpenAI addresses in this motion only the Plaintiffs in the Doe 3 Action (Case 3:22-cv-07074).

27 <sup>3</sup> Plaintiffs lump together GitHub, Microsoft, and OpenAI entities as “Defendants” throughout the  
28 complaint. (*See* Compl. at n.1.)

1 released Codex and Copilot—two “assistive AI-based systems” that are alleged to generate  
2 copied copyrighted material without attribution in some instances. (*Id.* ¶¶ 46, 77.) Plaintiffs  
3 allege that after Defendants trained Copilot and Codex using data gathered from publicly  
4 accessible repositories on GitHub, they used Copilot and Codex to distribute similar code to  
5 users. (*Id.* ¶¶ 46, 140.) In doing so, Plaintiffs contend that Defendants violated open-source  
6 licenses and infringed intellectual property rights. (*Id.* ¶¶ 143-171.)

7 Based on these allegations, Plaintiffs seek to represent two classes: an “Injunctive Relief  
8 Class” under Rule 23(b)(2) and a “Damages Class” under Rule 23(b)(3). (*Id.* ¶ 34.) Plaintiffs  
9 define both classes as “[a]ll persons or entities domiciled in the United States that, (1) owned an  
10 interest in at least one US copyright in any work; (2) offered that work under one of GitHub’s  
11 Suggested Licenses; and (3) stored Licensed Materials in any public GitHub repositories at any  
12 time” between January 1, 2015 and the present (the “Class Period”). (*Id.*)

13 The complaint asserts ten causes of action against the OpenAI Entities: (1) violation of the  
14 Digital Millennium Copyright Act (“DMCA”), 17 U.S.C. § 1202, *et seq.*; (2) breach of contract;  
15 (3) “tortious interference in a contract relationship”; (4) false designation of origin under the  
16 Lanham Act, 15 U.S.C. § 1125; (5) unjust enrichment under common law and California  
17 Business & Professions Code § 17200, *et seq.*; (6) violations of the California Consumer Privacy  
18 Act (“CCPA”), California Civil Code § 1798.150; (7) negligence; (8) unfair competition under  
19 common law, the Lanham Act, 15 U.S.C. § 1125, and California Business & Professions Code §  
20 17200, *et seq.*; (9) civil conspiracy; and (10) declaratory relief.<sup>4</sup>

### 21 **III. LEGAL STANDARD**

#### 22 **A. Motion to Dismiss for Lack of Subject Matter Jurisdiction.**

23 Federal courts “possess only that power authorized by Constitution and statute.”  
24 *Kokkonen v. Guardian life Ins. Co. of Am.*, 511 U.S. 375, 377 (1994). Article III “limits the  
25

---

26 <sup>4</sup> Where the complaint refers to common law or state law, OpenAI assumes for purposes of this  
27 motion that Plaintiffs have asserted such claims under California law. Although OpenAI does not  
28 concede that California law can be applied to acts occurring outside California here, even as pled  
under California law, Plaintiffs’ common law and state law claims should be dismissed.

jurisdiction of federal courts to ‘Cases’ and ‘Controversies.’” *Lance v. Coffman*, 549 U.S. 437, 439 (2007). If a plaintiff lacks Article III standing to bring a suit, the federal court lacks subject matter jurisdiction, and the suit must be dismissed under Rule 12(b)(1). *Cetacean Cmty. v. Bush*, 386 F.3d 1169, 1174 (9th Cir. 2004).

Moreover, where a complaint fails to disclose the identities of anonymous plaintiffs, in violation of Rule 10(a)’s requirement that the complaint “name all parties,” dismissal under Rule 12(b)(1) is appropriate. *Roe v. San Jose Unified Sch. Dist. Bd.*, No. 20-CV-02798-LHK, 2021 WL 292035, at \*10 (N.D. Cal. Jan. 28, 2021).

Once a defendant has moved to dismiss under Rule 12(b)(1), the plaintiff bears the burden of establishing the court’s jurisdiction. *See Chandler v. State Farm Mut. Auto Ins. Co.*, 598 F.3d 1115, 1122 (9th Cir. 2010).

**B. Motion to Dismiss for Failure to State a Claim Under Rule 12(b)(6).**

To satisfy Rule 8 and survive a Rule 12(b)(6) motion to dismiss, “a complaint must contain sufficient factual matter, accepted as true, to state a claim to relief that is plausible on its face.” *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (cleaned up). Dismissal is appropriate “where the complaint lacks a cognizable legal theory or sufficient facts to support a cognizable legal theory.” *Mendiondo v. Centinela Hosp. Med. Ctr.*, 521 F.3d 1097, 1104 (9th Cir. 2008). Where a plaintiff raises generalized allegations against multiple defendants, the complaint has not “stated sufficient facts to state a claim for relief that is plausible against *one* [d]efendant.” *In re iPhone App. Litig.*, No. 11-MD-022590-LHK, 2011 WL 4403963, at \*3 (N.D. Cal. Sept. 20, 2011). Moreover, “[f]actual allegations must be enough to raise a right to relief above the speculative level” and “a formulaic recitation of the elements of a cause of action will not do.” *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555 (2007). The Court need not accept as true conclusory allegations or legal characterizations, nor need it accept unreasonable inferences or unwarranted factual deductions. *In re Gilead Scis. Sec. Litig.*, 536 F.3d 1049, 1055 (9th Cir. 2008).



1 **IV. ARGUMENT**

2 **A. The Complaint Fails for Reasons Applicable to All Causes of Action.**

3 **1. Plaintiffs Lack Article III Standing to Assert Their Claims.**

4 The complaint must be dismissed because Plaintiffs have failed to sufficiently plead that  
5 they suffered a cognizable injury to satisfy “the irreducible constitutional minimum of standing”  
6 under Article III. *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992). The plaintiff bears  
7 the burden of establishing standing “for each claim [s]he seeks to press and for each form of relief  
8 that is sought.” *Davis v. FEC*, 554 U.S. 724, 734 (2008) (cleaned up). To satisfy Article III  
9 standing, a plaintiff must allege: (1) an injury in fact that is concrete and particularized, as well as  
10 actual or imminent; (2) that the injury is fairly traceable to the challenged action; and (3) that it is  
11 likely (not merely speculative) that injury will be redressed by a favorable decision. *Id.* at 733. A  
12 plaintiff does not satisfy the standing requirement “[w]hen speculative inferences are necessary . .  
13 . to establish [the] injury.” *Johnson v. Weinberger*, 851 F.2d 233, 235 (9th Cir. 1988) (cleaned  
14 up). In a putative class action, the named plaintiffs seeking to represent the class must establish  
15 that they *personally* have standing to bring the action. *See Lewis v. Casey*, 518 U.S. 343, 357  
16 (1996) (“[N]amed plaintiffs who represent a class must allege and show that they personally have  
17 been injured, not that injury has been suffered by other, unidentified members of the class to  
18 which they belong and which they purport to represent.”) (cleaned up); *Birdsong v. Apple, Inc.*,  
19 590 F.3d 955, 960 (9th Cir. 2009) (affirming dismissal of putative class action brought by iPod  
20 users for lack of standing where “[t]he risk of injury the plaintiffs allege is not concrete and  
21 particularized *as to themselves*”); *see also Lierboe v. State Farm Mut. Auto. Ins. Co.*, 350 F.3d  
22 1018, 1022-23 (9th Cir. 2003) (vacating class certification where named plaintiff lacked standing  
23 to assert a claim under state law).

24 Here, the complaint contains no allegation or explanation of whether and how any  
25 Plaintiff was harmed. Plaintiffs rely entirely on generic descriptions of the alleged practices of  
26 the OpenAI Entities to support their theory of injury. (*See, e.g.,* Compl. ¶¶ 90-91.) Plaintiffs do  
27 not allege that Copilot or Codex reproduced *their* code or disclosed *their* personal information.  
28 Instead, the complaint describes the purported reproduction of the code or personal information of



others. (See Compl. ¶¶ 48-63, 68-77, 87-89.) Plaintiffs have not provided a single example nor alleged any injury that is concrete and particularized *as to them*. This is insufficient under Article III. See, e.g., *Alsheikh v. Lew*, No. 3:15-CV-03601-JST, 2016 WL 1394338, at \*2-3 (N.D. Cal. Apr. 7, 2016) (dismissing claim for lack of Article III standing where plaintiff did not “identif[y] any particular injury that he has suffered”).

## 2. Plaintiffs Have Failed to Obtain Leave to Proceed Anonymously.

The complaint must be dismissed under Rule 12(b)(1) because Plaintiffs failed to comply with Federal Rule of Civil Procedure 10(a), which requires that a “complaint name all the parties.” This rule reflects the “paramount importance of open courts.” *Doe v. Kamehameha Schs./Bernice Pauahi Bishop Est.*, 596 F.3d 1036, 1046 (9th Cir. 2010) (affirming dismissal of complaint based on plaintiffs’ failure to disclose identities). Plaintiffs have not identified themselves. And while, in the Ninth Circuit, they may seek leave to proceed anonymously after filing the complaint, they have not done so and could not succeed were they to try. See, e.g., *Doe v. UNUM*, 164 F. Supp. 3d 1140, 1144 (N.D. Cal. 2016) (dismissing complaint). A party may proceed anonymously only where “special circumstances justify secrecy.” *Does 1 Thru XXIII v. Advanced Textile Corp.*, 214 F.3d 1058, 1067 (9th Cir. 2000). Courts “must balance the need for anonymity against the general presumption that parties’ identities are public information and the risk of unfairness to the opposing party.” *Id.* at 1068. Under this balancing test, the Ninth Circuit has identified three situations in which parties may proceed anonymously: (1) “when identification creates a risk of retaliatory physical or mental harm;” (2) “when anonymity is necessary to ‘preserve privacy in a matter of a sensitive and highly personal nature;” or (3) “when the anonymous party is ‘compelled to admit [his or her] intention to engage in illegal conduct, thereby risking criminal prosecution.’” *Id.* at 1068 (citations omitted). Plaintiffs provide no facts at all justifying their request to keep their identities hidden and provide no examples of Plaintiffs’ code they claim to be at issue, and thus make it impossible for OpenAI to respond to their allegations. Accordingly, unless and until Plaintiffs are willing to put their names and their code

on the allegations they have made, dismissal is appropriate.<sup>5</sup>

### 3. The Complaint’s Undifferentiated Allegations Against the Six OpenAI Entities Fail to Satisfy Pleading Requirements.

Plaintiffs have also failed to specify what acts they contend each of the six OpenAI Entities committed individually, requiring dismissal. A complaint that “lumps together multiple defendants in one broad allegation fails to satisfy the notice requirement of Rule 8(a)(2).”

*Sebastian Brown Prods. v. Muzooka Inc.*, 143 F. Supp. 3d 1026, 1037 (N.D. Cal. 2015) (cleaned up). Where a plaintiff sues multiple defendants, “the complaint must specify exactly what each separate defendant is alleged to have done to cause plaintiff harm.” *Rosas v. City of Santa Rosa*, No. 21-CV-06179-JST, 2022 WL 2158968, at \*2 (N.D. Cal. June 15, 2022) (cleaned up).

Plaintiffs fail to meet this standard. Through the body of the complaint, the six OpenAI Entities are referred to collectively as “OpenAI.” (*See, e.g.*, Compl. at 1 n.1.) After alleging that “OpenAI” was involved in “launch[ing] Copilot” and “debut[ing] its Codex product,” the complaint proceeds to accuse the various OpenAI Entities collectively of violating Plaintiffs’ “intellectual-property rights, licenses, and other rights” by using unidentified code from GitHub’s public repositories. (*See id.* ¶¶ 8-9, 139.) But this is not a case in which the acts alleged against each OpenAI Entity are the same. Indeed, the complaint concedes that each OpenAI Entity performed distinct activities and that not all Entities played a role in developing Copilot and Codex. (*Compare id.* ¶¶ 23-28.) In describing the parties, Plaintiffs allege that OpenAI, Inc. and OpenAI, L.P. “programmed, trained, and maintains Codex,” which provides “an integral piece of Copilot.” (*Id.* ¶ 23.) The complaint does not allege that OpenAI GP, L.L.C., OpenAI Startup Fund I, L.P., OpenAI Startup Fund GP I, L.L.C., and OpenAI Startup Fund Management, LLC played any role in developing Codex or Copilot. (*See id.* ¶¶ 25-28.)

Plaintiffs’ indiscriminate lumping together of the OpenAI Entities violates Rule 8(a)(2)’s

---

<sup>5</sup> The Defendants notified Plaintiffs that their complaint was procedurally deficient under Rule 10(a) on January 11, 2023. (*See* Declaration of Michael Jacobs, filed herewith, Ex. 1). Plaintiffs responded that they would seek approval to proceed anonymously but have yet to file the requisite motion. (*See id.* Ex. 2.)

notice requirement. The complaint fails to provide fair notice to Defendants as it merely alleges wrongdoing against “OpenAI” or “Defendants” without specifying which OpenAI Entity allegedly committed the act. (*See* Compl. ¶¶ 8-9, 14, 37, 43-44, 47, 63, 66, 82-84, 134-141.)

**B. The Copyright Act Preempts Several State Law Causes of Action.**

Federal law preempts Plaintiffs’ claims for tortious interference in a contractual relationship, unjust enrichment, and unfair competition, and accordingly, provides another basis for dismissal. Preemption under Section 301 of the Copyright Act applies if (1) “the ‘subject matter’ of the state law claim falls within the subject matter of copyright as described in 17 U.S.C. §§ 102 and 103” and (2) “whether the rights asserted under state law are equivalent to the rights contained in 17 U.S.C. § 106, which articulates the exclusive rights of copyright holders.” *Maloney v. T3Media, Inc.*, 853 F.3d 1004, 1010 (9th Cir. 2017) (cleaned up). Courts have concluded that the Copyright Act precludes the following:

- **Tortious interference with a contract.** Plaintiffs claim that the OpenAI Entities have “wrongfully interfered with the business interests and expectations of Plaintiffs ... by improperly using Copilot to create Derivative Works that compete against” Plaintiffs’ works. (Compl. ¶ 189.) This, in essence, boils down to an allegation that is “not qualitatively different from [a] copyright infringement” claim. *See Media.net Advert. FZ-LLC v. NetSeer, Inc.*, 156 F. Supp. 3d 1052, 1072 (N.D. Cal. 2016) (concluding preemption where plaintiff alleged that defendant directly copied plaintiff’s code to create its own product and “undermined [p]laintiff’s relationship with Microsoft by representing that [its] products could work just as well”).

- **Unjust enrichment.** The Copyright Act preempts Plaintiffs’ unjust enrichment claim because the crux of this claim asserts that OpenAI improperly benefitted from using Licensed Materials to create Derivative Works. *Del Madera Props. V. Rhodes & Gardner, Inc.*, 820 F.2d 973, 977 (9th Cir. 1987) (finding preemption because an “implied promise not to use or copy materials within the subject matter of copyright is equivalent” to the Copyright Act’s protections); *Firoozye v. Earthlink Network*, 153 F. Supp. 2d 1115, 1126 (N.D. Cal. 2001) (finding preemption where plaintiff alleged defendant improperly benefitted from using copyrighted software and “that a contract should be implied in law (e.g., a quasi-contract [] or unjust enrichment claim)”).



• **Unfair Competition.** To the extent Plaintiffs’ claims are based on preempted state law claims, the derivative claim must also fail. *See Sulit v. Sound Choice Inc.*, No. C06-00045 MJJ, 2006 WL 8442163, at \*7 (N.D. Cal. Nov. 14, 2006) (“State law causes of action for unfair competition based on misappropriation of copyrighted material are preempted,” but, where the state claim acts as “a tort of ‘passing off,’ it is not preempted.”).

**C. Plaintiffs’ Claims Fail for Reasons Specific to Each Claim.**

**1. Plaintiffs’ DMCA Claim Should be Dismissed.**

Although the complaint is replete with allegations about alleged similarities between Copilot’s output and the code it was trained on, Plaintiffs do not assert a copyright infringement claim. Instead, they allege that Defendants violated the DMCA by (1) removing or altering CMI from Licensed Materials, (2) distributing copies of Licensed Materials knowing CMI had been removed or altered without authority, and (3) knowingly providing CMI that is false by “asserting and/or implying that Copilot is the author of the Licensed Materials.” (Compl. ¶¶ 158-159.) Plaintiffs’ allegations do not meet DMCA requirements and fail properly to plead a DMCA claim.

**a. Plaintiffs Have Not Properly Pled a Claim for Removal of CMI.**

To properly plead a claim for removal of CMI, a plaintiff must plausibly allege: (1) the existence of CMI on a work, (2) removal or alteration of that information, (3) that the removal or alteration was done intentionally; and (4) the removal or alteration was done knowing or having reasonable grounds to know that it would induce, enable, facilitate, or conceal copyright infringement. 17 U.S.C. § 1202(b); *Stevens v. CoreLogic, Inc.*, 899 F.3d 666, 673 (9th Cir. 2018) (discussing the mental state elements); *O’Neal v. Sideshow, Inc.*, 583 F. Supp. 3d 1282, 1286-87 (C.D. Cal. 2022) (discussing other elements).

**(i) Failure to Allege Removal from Identical Copies.**

Plaintiffs’ claim under § 1202(b) arises out of the allegation that CMI was removed from Plaintiffs’ code. But in order to prevent § 1202 from subsuming every copyright dispute, courts have interpreted “removal” in the § 1202 context to require that there was some *identical* copy of the plaintiff’s work made without the plaintiff’s CMI. *See, e.g., Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) (requiring that CMI was removed from “a plaintiff’s



product or original work”), *aff’d and rev’d in part on other grounds*, 336 F.3d 811 (9th Cir. 2003). Where a defendant makes a copy of a defendant’s work that is substantially similar, but not identical, to the plaintiff’s work, and omits CMI from that copy, there may be a claim for copyright infringement, but there cannot be a claim under § 1202. *See Frost-Tsuji Architects v. Highway Inn, Inc.*, No. CIV. 13-00496 SOM, 2015 WL 263556, at \*3 (D. Haw. Jan. 21, 2015), *aff’d*, 700 F. App’x 674 (9th Cir. 2017) (“But the drawing by [the defendant] is not identical to the drawing by [the plaintiff], such that this court can say that [the defendant] removed or altered [the plaintiff’s] copyright management information from [the drawing].”); *id.* (“basing a drawing on [the plaintiff’s] work is not sufficient to support a claim” under § 1202); *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at \*6 (C.D. Cal. Aug. 14, 2020) (dismissing DMCA claim because “while the works may be *substantially similar*, Defendant did not make *identical* copies of Plaintiff’s works and then remove engraved CMI”).

Here, Plaintiffs concede that Copilot does not generate identical copies:

- “[T]he Output is often a *near-identical* reproduction of code from the training data.” (Compl. ¶ 46 (emphasis added));
- “Codex has reproduced Haverbeke’s Licensed Material *almost verbatim*, with the only difference being drawn from a different portion of those same Licensed Materials.” (*Id.* ¶ 60 (emphasis added));
- “Like the other examples above—and most of Copilot’s Output—this output is *nearly a verbatim copy* of copyrighted code. In this case, it is *substantially similar* to the “isPrime” function in the book *Think JavaScript* by Max X. Curinga et al., . . . .” (*Id.* ¶ 74 (emphasis added).)

Because Plaintiffs affirmatively allege that the output at issue is not identical to the allegedly copied material, they have pled themselves out of court on the § 1202 claim, and it should be dismissed with prejudice.

## (ii) Failure to Identify the Works.

The § 1202 claim is also subject to dismissal because Plaintiffs have not sufficiently identified any works from which CMI was allegedly removed. *See Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1175 (N.D. Cal. 2019) (dismissing DMCA claim because Menzel

1 “merely alleged that his photographs ‘were altered to remove certain of [his] copyright  
2 management information’ without providing any facts to identify which photographs had CMI  
3 removed or to describe what the removed or altered CMI was”). Plaintiffs merely allege  
4 generally that Defendants removed CMI from “Licensed Materials,” which they define broadly as  
5 “materials made available publicly on GitHub that are subject to various licenses containing  
6 conditions for use of those works.” (Compl. ¶¶ 1, 148.) The complaint highlights Plaintiffs’  
7 imprecision, reciting in the DMCA cause of action that Copilot was “trained on millions—  
8 possibly billions—of lines of code.” (*Id.* ¶ 143.) But the few specific instances the complaint  
9 points to are not examples of Plaintiffs’ own code, but snippets from third-party programming  
10 textbooks. (*Id.* ¶¶ 56-61, 71-75.) That code is not the subject of Plaintiffs’ DMCA allegations,  
11 as it does not fall within the complaint’s definition of “Licensed Materials.” Without identifying  
12 specific works from which CMI was removed, Plaintiffs fail to state a claim for CMI removal.

13 **(iii) Failure to Adequately Plead Scienter.**

14 Plaintiffs also have not pled facts sufficient to meet the “double-scienter” requirement of  
15 Section 1202(b)(3), which requires “the defendant who distributed improperly attributed  
16 copyrighted material must have actual knowledge that CMI ‘has been removed or altered without  
17 authority of the copyright owner or the law,’ as well as actual or constructive knowledge that such  
18 distribution ‘will induce, enable, facilitate, or conceal an infringement.’” *Mango v. BuzzFeed,*  
19 *Inc.*, 970 F.3d 167, 171 (2d Cir. 2020). “[T]he plaintiff must provide evidence from which one  
20 can infer that future infringement is likely, albeit not certain, to occur as a result of the removal or  
21 alteration of CMI.” *CoreLogic*, 899 F.3d at 676 (finding CoreLogic not liable for violating  
22 § 1202(b) because photographers had “not put forward *any* evidence that CoreLogic knew its  
23 software carried even a substantial risk of inducing, enabling, facilitating, or concealing  
24 infringement, let alone a pattern or probability of such connection to infringement”).

25 Here, Plaintiffs have not alleged facts sufficient to establish a substantial risk that any  
26 copyright infringement has occurred or that any future infringement is likely because of the  
27 removal of CMI, nor that any of the OpenAI Entities had reason to know of any such  
28 likelihood. They have not alleged, for example, copying of protectible expression: that is, that the



allegedly copied code was original, that there was no merger of idea and expression, and that the allegedly copied code did not represent “*scènes à faire*.” See, e.g., *Oracle Am., Inc. v. Google Inc.*, 872 F. Supp. 2d 974, 984-997 (N.D. Cal. 2012) (summarizing the various doctrines limiting copyright in computer programs). And they would need to allege that any copying was not fair use—a heavy burden in light of the Supreme Court’s holding in the source-code context that “taking only what was needed to allow users to put their accrued talents to work in a new and transformative program . . . was a fair use of that material as a matter of law.” *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1209 (2021); see also, e.g., *Authors Guild v. Google, Inc.*, 804 F.3d 202, 225 (2d Cir. 2015) (copying of millions of books for the purpose of searching them and providing relevant snippets to users was fair use). Finally, they would have to identify with specificity which work or works were copied and specify which defendant is alleged to have infringed which particular copyright. *Lynwood Invs. CY Ltd. v. Konovalov*, No. 20-CV-03778-MMC, 2022 WL 3370795, at \*19 (N.D. Cal. Aug. 16, 2022) (dismissing claim). All of these are substantial hurdles to showing that Defendants had reason to know that they would cause or further copyright infringement. The complaint meets none of them.

**b. Plaintiffs Have Failed to Plead a Claim for Distributing Copies of Works from Which CMI Has Been Removed.**

Plaintiffs’ claim that Defendants have distributed copies of code from which CMI has been removed fails for the same reasons as its claim for removal of CMI. 17 U.S.C. §§ 1202(b)(2), 1202(b)(3). See *Kirk Kara*, 2020 WL 5991503, at \*6 (applying same 1202(b)(1) analysis to distribution claims); *Dolls Kill, Inc. v. Zoetop Bus. Co.*, No. 2:22-cv-01463-RGK-MAA, 2022 WL 16961477, at \*3-4 (C.D. Cal. Aug. 25, 2022) (concluding no DMCA violation for complaint that defendants “are distributing knockoff products” where the works were not identical and only had “certain[] similarities”); *Mango v. BuzzFeed, Inc.*, 356 F. Supp. 3d 368, 376 (S.D.N.Y 2019), *aff’d*, 970 F.3d 167 (2d Cir. 2020) (in view of few decisions involving DMCA’s distribution prohibitions, looking to CMI removal caselaw for guidance). Plaintiffs have not specifically identified any such copies; the supposed copies are not identical; and Plaintiffs have not shown the requisite scienter.

**c. Plaintiffs Have Failed to Show that OpenAI Has Conveyed Any False CMI in Connection with Copilot Outputs.**

Plaintiffs' claim that Defendants have conveyed false CMI (Compl. ¶¶ 158-159) is also fundamentally flawed. 17 U.S.C. § 1202(a). The DMCA defines CMI as any information identifying the work, its author or copyright owner, and the terms and condition of use, or "links to such information," "conveyed in connection with copies . . . of [the] work." 17 U.S.C. § 1202(c) (emphasis added). Courts require that the allegedly false CMI's location suggest an association with plaintiff's work. *See SellPoolSuppliesOnline.com, LLC v. Ugly Pools Ariz., Inc.*, 804 F. App'x 668, 670-71 (9th Cir. 2020) (affirming grant of summary judgment against plaintiff's false CMI claim, finding defendant's copyright notice at the bottom of the webpage was not "conveyed in connection with" plaintiff's photos); *Logan v. Meta Platforms, Inc.*, No. 22-cv-1847-CRB, 2022 WL 14813836, at \*8 (N.D. Cal. Oct. 25, 2022) (finding copyright notice on the bottom of each Facebook user page separated from the rest of the content insufficient to plead that Meta conveyed CMI in connection with plaintiff's photos).

Once again, Plaintiffs have not pointed to any specific code that Defendants conveyed containing false CMI, as pleading standards require. *See* § IV.C.1.a, *supra*.

Moreover, Plaintiffs have not properly alleged that the OpenAI Entities have conveyed any CMI at all. Plaintiffs merely allege that "Defendants have a business practice of asserting and/or implying that Copilot is the author of the Licensed Materials." (Compl. ¶ 158.) But that is not the same as conveying CMI "in connection with copies" of the work. And in the complaint's examples of Codex-generated code, there is no CMI presented whatsoever. (*See id.* ¶¶ 49, 69.)

**2. Plaintiffs' Breach of Contract Claim Fails.**

To state a claim for breach of contract under California law, a plaintiff must "plead the contract, plaintiff's performance (or excuse for nonperformance), defendant's breach, and damage to plaintiff therefrom." *Low v. LinkedIn Corp.*, 900 F. Supp. 2d 1010, 1028 (N.D. Cal. 2012) (cleaned up). When asserting written contract claims, a plaintiff must "allege the specific provisions in the contract creating the obligation the defendant is said to have breached." *Young v. Facebook, Inc.*, 790 F. Supp. 2d 1110, 1117 (N.D. Cal. 2011). Here, Plaintiffs have not



adequately pled the existence of a contract between themselves and any of the OpenAI Entities or sufficient factual detail supporting that OpenAI Entities allegedly breached any such agreement.

**a. Plaintiffs Have Not Sufficiently Pled Existence of a Contract.**

Plaintiffs have not adequately pled facts demonstrating the existence of a contract between Plaintiffs and any OpenAI Entity. *See CSI Elec. Contractors, Inc. v. Zimmer Am. Corp.*, No. CV 12-10876-CAS (AJWx), 2013 WL 1249021, at \*3 (C.D. Cal. Mar. 25, 2013) (granting motion to dismiss breach of contract claim because complaint failed to adequately allege existence of a contract). Here, the complaint merely alleges that “Plaintiffs . . . offer code under various [unspecified] Licenses” and attaches a sampling of the most common licenses in an appendix. (Compl. ¶ 173, App. A.) The complaint then alleges that contracts have been formed with Defendants, collectively, based on their use of code subject to certain licenses. (*Id.* ¶ 175.) However, the complaint does not indicate which licenses are at issue or which provisions the OpenAI Entities allegedly breached. These vague and conclusory allegations are insufficient to establish the existence of a contract. *See Ramirez v. GMAC Mortg.*, No. CV 09-8189 PSG (FFMx), 2010 WL 148167, at \*2 (C.D. Cal. Jan. 12, 2010) (holding plaintiff failed to plead existence of a contract where it did not “set forth [the] terms of the contract that Defendants’ conduct is alleged to have breached”).

**b. Plaintiffs Fail to Allege Facts Demonstrating the Contractual Provisions OpenAI Entities Allegedly Breached.**

While Plaintiffs allege generally that OpenAI breached open source licenses by failing to (i) provide attribution, (ii) include copyright notices, and (iii) identify the license applicable to the work (Compl. ¶¶ 181-183), the complaint fails to identify the particular terms of the alleged agreements that the OpenAI Entities purportedly violated. *See, e.g., Zepeda v. PayPal, Inc.*, 777 F. Supp. 2d 1215, 1220 (N.D. Cal. 2011) (dismissing contract claim in part because plaintiffs failed to identify any provision in PayPal’s user agreement which prohibited PayPal’s conduct).

“Facts alleging a breach, like all essential elements of a breach of contract cause of action, must be pleaded with specificity.” *Rubio v. U.S. Bank N.A.*, No. C 13-05752 LB, 2014 WL 1318631, at \*9 (N.D. Cal. Apr. 1, 2014) (cleaned up); *see also Sutherland v. Francis*, No. 12-CV-

05110-LHK, 2014 WL 879697, at \*4 (N.D. Cal. Mar. 3, 2014) (dismissing contract claim that did not include “the essential terms of the agreement and more specific allegations as to breach”). Plaintiffs’ generic allegations fail to satisfy this requirement.

### 3. The Claim for Tortious Interference in Contractual Relationship Fails.

Plaintiffs have not alleged facts supporting the existence of a contract between Plaintiffs and a third party or the requisite intent to disrupt it. A tortious interference with contract claim requires: “(1) a valid contract between plaintiff and a third party; (2) defendant’s knowledge of this contract; (3) defendant’s intentional acts designed to induce a breach or disruption of the contractual relationship; (4) actual breach or disruption of the contractual relationship; and (5) resulting damage.” *Piping Rock Partners, Inc. v. David Lerner Assocs., Inc.*, 946 F. Supp. 2d 957, 979 (N.D. Cal. 2013) (cleaned up). “The mere allegation that [defendant] ‘purposefully and intentionally interfered’ with a contract, without any factual support ... does not satisfy the requirements for stating a claim for tortious interference with contractual relations.” *Wynn v. NBC*, 234 F. Supp. 2d 1067, 1122 (C.D. Cal. 2002).

The complaint alleges that “Defendants have wrongfully interfered with the business interests and expectations of Plaintiffs . . . by improperly using Copilot to create Derivative Works that compete against OSC.” (Compl. ¶ 189.) But the complaint does not specify the “business interest and expectations” with which the OpenAI Entities allegedly interfered. At most, the complaint alleges, albeit in connection with other causes of action, that the OpenAI Entities have interfered with Plaintiffs’ “contractual relationship with users of their code.” (*Id.* ¶¶ 212(b), 244(b).) However, these vague allegations do not identify a specific contract or contractual provision between Plaintiffs and the unidentified users. Absent a valid contract and contractual provisions that the OpenAI Entities caused others to breach, Plaintiffs’ tortious interference claim fails. *See Image Online Design, Inc. v. Internet Corp. for Assigned Names & Numbers*, No. CV 12-08968 DDP (JCx), 2013 WL 489899, at \*9 (C.D. Cal. Feb. 7, 2013) (dismissing interference claim where plaintiff “has not alleged any facts identifying the particular contracts, the actual disruption of these contracts, or any actual damage to [plaintiff],” and it “cannot simply allege that [defendant] has interfered with its business model”).



1 Plaintiffs also do not allege any intentional actions undertaken by the OpenAI Entities that  
2 were intended to induce the purported users of Plaintiffs' code to breach their agreement with  
3 Plaintiffs. To satisfy the intent element, a plaintiff must allege: (1) that defendant specifically  
4 intended to disrupt the relationship, or (2) that the defendant knew that the interference was  
5 certain or substantially certain to occur as a result of its action. *Korea Supply Co. v. Lockheed*  
6 *Martin Corp.*, 29 Cal. 4th 1134, 1154 (2003). Here, Plaintiffs' allegations fall short of making  
7 the necessary showing of intent for two reasons.

8 *First*, Plaintiffs do not allege any particular act by the OpenAI Entities designed to disrupt  
9 Plaintiffs' alleged relationship with users of their code. For example, there is no allegation that  
10 the OpenAI Entities contacted the users or otherwise tried to induce them to breach their  
11 purported contract with Plaintiffs. Plaintiffs' failure to allege specific intent requires dismissal.  
12 *See name.space, Inc. v. Internet Corp. for Assigned Names & Numbers*, No. CV 12-8676 PA  
13 (PLAx), 2013 WL 2151478, at \*8 (C.D. Cal. Mar. 4, 2013) (dismissing tortious interference  
14 claim because "the Complaint does not allege any intentional actions undertaken by [defendant]  
15 designed to induce breach of Plaintiff's contracts with its clients or any evidentiary facts, as  
16 opposed to conclusory allegations, of actual breach or disruption and resulting damage").

17 *Second*, Plaintiffs do not allege that the OpenAI Entities were substantially certain that  
18 their actions would disrupt Plaintiffs' alleged relationship with users of their code. A plaintiff  
19 must show "an interference that is incidental to the actor's independent purpose and desire but  
20 known to him to be a necessary consequence of his action." *Korea Supply*, 29 Cal. 4th at 1155-56  
21 (cleaned up). Nowhere do Plaintiffs allege that the OpenAI Entities knew the "interference" was  
22 certain or substantially certain to occur because of their actions.

#### 23 **4. Plaintiffs Fail to Allege a False Designation of Origin Claim.**

24 A claim for false designation of origin must relate to the origin of tangible goods, not the  
25 authorship of an intangible work like computer code. 15 U.S.C. § 1125(a)(1)(A); *Dastar Corp. v.*  
26 *Twentieth Century Fox Film Corp.*, 539 U.S. 23, 37 (2003) (concluding that the phrase "origin of  
27 goods...refers to the producer of the tangible goods that are offered for sale, and not to the author  
28 of any idea, concept or communication embedded in those goods"); *Agence France Presse v.*

1 *Morel*, 769 F. Supp. 2d 295, 307 (S.D.N.Y. 2011) (holding that *Dastar* forecloses Lanham Act  
2 claims relating to authorship). To hold that authorship is actionable under the Lanham Act would  
3 “provide authors of creative works with perpetual protection under the Lanham Act that they did  
4 not have under the Copyright Act.” *Oppenheimer v. Allvoices, Inc.*, No. C 14-00499 LB, 2014  
5 WL 2604033, at \*11 n.10 (N.D. Cal. June 10, 2014).

6 Plaintiffs’ claim is precisely the kind of false designation of origin claim foreclosed by  
7 these precedents. The complaint alleges that “GitHub and OpenAI” have passed off the code  
8 contained in Copilot’s output as originating from Copilot, GitHub, and/or OpenAI, thereby  
9 violating the Lanham Act. (Compl. ¶¶ 201, 212.) Plaintiffs also claim that “Codex does not  
10 identify the owner of the copyright to [its] Output” because “it has not been trained to provide  
11 Attribution” and that “[a]s with Codex, Copilot does not provide the end user any attribution of  
12 the original author of the code, nor anything about their license requirements.” (*Id.* ¶¶ 56, 77.)  
13 Even taking these allegations as true, the complaint does not give rise to a Lanham Act claim  
14 because the alleged misrepresentation relates to authorship.

##### 15 **5. Plaintiffs Fail to State a Claim for Unjust Enrichment.**

16 Plaintiffs’ claim for unjust enrichment fails because it is not an independent cause of  
17 action. “[I]n California, there is no[] standalone cause of action for ‘unjust enrichment.’”  
18 *Astiana v. Hain Celestial Grp., Inc.*, 783 F.3d 753, 762 (9th Cir. 2015). Courts instead “construe  
19 [an unjust enrichment claim] as a quasi-contract claim for restitution.” *Id.* (cleaned up). But  
20 plaintiffs cannot recover on a quasi-contract claim if they also seek to recover under a breach of  
21 contract claim, as they do here. *See Klein v. Chevron U.S.A., Inc.*, 202 Cal. App. 4th 1342, 1388  
22 (2012) (A plaintiff may not “recover on a quasi-contract claim if the parties have an enforceable  
23 agreement.”). Because Plaintiffs do not allege the absence of an enforceable contract—and in  
24 fact allege the opposite—Plaintiffs’ unjust enrichment claim should be dismissed. In any event,  
25 Plaintiffs have not pled the required elements for unjust enrichment as they have not adequately  
26 alleged that the OpenAI Entities “ha[ve] been unjustly conferred a benefit through mistake, fraud,  
27 coercion, or request.” *See Astiana*, 783 F.3d at 762. (cleaned up). Finally, to the extent  
28 Plaintiffs’ “unjust enrichment claim” under the UCL is seeking restitution based on an alleged



violation of the UCL, that claim fails for the same reasons discussed in Section B.6.below.

# **6. Plaintiffs Fail to State an Unfair Competition Claim.**

Plaintiffs assert an unfair competition claim under (1) the Lanham Act, (2) California’s UCL statute, and (3) common law, all of which are predicated on the OpenAI Entities’ alleged violations of the DMCA, tortious interference with contract relations, false designation of origin, and violations of the CCPA and California’s constitutional right to privacy. (Compl. ¶ 212.) Under any theory, plaintiffs have failed to state a claim.

Plaintiffs’ UCL claim, brought only under the “unlawful” prong, fails because there is no predicate violation. When the underlying legal claim that supports a UCL cause of action fails, “so too will the [] derivative UCL claim.” *Yellowcake, Inc. v. Hyphy Music, Inc.*, No. 1:20-CV-0988 AWI BAM, 2021 WL 3052535, at \*13 (E.D. Cal. July 20, 2021); *see also Eidmann v. Walgreen Co.*, 522 F. Supp. 3d 634, 647 (N.D. Cal. 2021) (If the “plaintiff cannot state a claim under the predicate law ... [the UCL] claim also fails.”) (cleaned up). For reasons discussed in this motion, plaintiffs have not adequately alleged any underlying legal claim.

Plaintiffs’ UCL claim separately fails because plaintiffs have not established that they lack an adequate legal remedy. “Remedies under the UCL are limited to restitution and injunctive relief, and do not include damages.” *Silvercrest Realty, Inc. v. Great Am. E&S Ins. Co.*, No. SACV 11-01197-CJC (ANx), 2012 WL 13028094, at \*2 (C.D. Cal. Apr. 4, 2012). To state a viable claim for “equitable restitution for past harm under the UCL,” a plaintiff “must establish that she lacks an adequate remedy at law.” *Sonner v. Premier Nutrition Corp.*, 971 F.3d 834, 844 (9th Cir. 2020) (affirming dismissal where plaintiff failed to allege an inadequate legal remedy). Here, Plaintiffs have not shown that no adequate legal remedy exists.

Plaintiffs also cannot identify any common law or Lanham Act basis for an unfair competition claim separate from their false designation of origin claim. In California, “[t]he common law tort of unfair competition is generally thought to be synonymous with the act of ‘passing off’ one’s goods as those of another.” *Sybersound Recs., Inc. v. UAV Corp.*, 517 F.3d 1137, 1153 (9th Cir. 2008) (explaining the tort provided “an equitable remedy against the wrongful exploitation of trade names and common law trademarks that were not otherwise

entitled to legal protection” and expansion of unfair competition law is primarily based in statute) (cleaned up). The Ninth Circuit “has consistently held that state common law claims of unfair competition and actions pursuant to [the UCL] are ‘substantially congruent’ to claims made under the Lanham Act.” *Sebastian Brown Prods. LLC v. Muzooka Inc.*, No. 15-CV-01720-LHK, 2016 WL 949004, at \*15 (N.D. Cal. Mar. 14, 2016) (cleaned up). Plaintiffs’ common law unfair competition claim therefore also fails because it is premised on the same conduct underlying their deficient false designation of origin claim. (*See* § IV.C.4, *supra*.)

#### 7. Plaintiffs Fail to Adequately Plead A Violation of the CCPA.

Plaintiffs’ claim under the CCPA suffers from numerous pleading defects. Plaintiffs lack statutory standing, and they failed to provide the required written notice to the OpenAI Entities prior to filing this complaint. Even if Plaintiffs satisfied these threshold requirements, the CCPA’s limited private right of action does not cover the alleged conduct. To the extent any of these allegations are actionable, Plaintiffs also failed to allege facts showing that the OpenAI Entities disclosed personal information protected by the CCPA.

As an initial matter, the named Plaintiffs cannot satisfy the statutory requirements for a CCPA claim. Out-of-state plaintiffs lack standing to assert a claim under the CCPA. Cal. Civ. Code § 1798.140(i) (defining “consumer” under the CCPA as “a natural person who is a California resident”); *see Hayden v. Retail Equation, Inc.*, No. SACV2001203DOCDFM, 2022 WL 2254461, at \*5 (C.D. Cal. May 4, 2022) (concluding that “CCPA claims brought by out-of-state Plaintiffs [] fail because the CCPA does not apply to non-California residents”). Here, Doe 3 and Doe 4 are not California residents. (*See* Compl. ¶¶ 19-20.)

Moreover, Plaintiffs failed to comply with the requirement to provide the OpenAI Entities with written notice of the alleged CCPA violations prior to filing the complaint. Section 1798.150(b) states that “*prior to initiating any action*” under the CCPA, the customer must provide “written notice identifying the specific provisions [that] have been or are being violated.” Cal. Civ. Code § 1798.150(b). No OpenAI Entity received such notice, and the complaint does not allege otherwise.

The CCPA claim also should be dismissed because it provides only a limited private right



of action: a consumer must allege that their “nonencrypted and nonredacted personal information” was “subject to an unauthorized access and exfiltration, theft, or disclosure as a result of a business’s violation of the duty to implement and maintain reasonable security procedures and practices appropriate to the nature of the information.” Cal. Civ. Code § 1798.150(a)(1). No cause of action exists for “violations of any other section of [the CCPA].” Cal. Civ. Code § 1798.150(c). It is well-settled that “a statute creates a private right of action only if the statutory language or legislative history affirmatively indicates such an intent.” *Lil’ Man In the Boat, Inc. v. City and Cnty. of San Francisco*, No. C17-CV-00904-JST, 2018 WL 4207260, at \*3 (N.D. Cal. Sept. 4, 2018) (cleaned up). The OpenAI Entities’ alleged failures to provide (i) an opt-out notice, (ii) a clear and conspicuous opt-out link, (iii) a right to deletion, and (iv) a right to access personal information, are not actionable because the CCPA unambiguously reserved enforcement of these provisions to the California Privacy Protection Agency. *See* Cal. Civ. Code § 1798.155(a).

Finally, Plaintiffs’ vague and conclusory allegations regarding the OpenAI Entities’ security practices fail to state a claim under the CCPA. “[P]lausibility pleading standards are especially important in cases like this, where the Defendant faces the ‘potentially enormous expense of discovery’ if the Court denies [a] motion to dismiss.” *Razuki v. Caliber Home Loans, Inc.*, No. 17cv1718-LAB (WVG), 2018 WL 6018361, at \*2 (S.D. Cal. Nov. 15, 2018) (dismissing data breach claims). In the absence of allegations that show the OpenAI Entities’ security procedures and practices were deficient, Plaintiffs cannot state a CCPA claim. *See Maag v. U.S. Bank Nat’l Ass’n*, No. 21-cv-00031-H-LL, 2021 WL 5605278, at \*2 (S.D. Cal. Apr. 8, 2021) (dismissing CCPA claim where plaintiff made unsupported allegations that “his PII was compromised because Defendant did not ‘implement and maintain security procedures and practices’ [and] ‘failed to effectively monitor its systems for security vulnerabilities’”); *Anderson v. Kimpton Hotel & Rest. Grp., LLC*, No. 19-CV-01860-MMC, 2019 WL 3753308, at \*3-4 (N.D. Cal. Aug. 8, 2019) (dismissing complaint that was “devoid of facts” regarding the “inadequate securities measures” that purportedly caused plaintiff’s alleged injury).

Here, Plaintiffs merely assert that the OpenAI Entities violated the CCPA by “collecting,

maintaining, and controlling their customers’ sensitive personal information” and “engineering, designing, maintaining, and controlling systems that exposed their customers’ sensitive personal information” without specifying what type of “sensitive personal information” had been collected or exposed. (Compl. ¶ 233.) Plaintiffs further allege that the OpenAI Entities were “aware of Copilot’s propensity for revealing PII” and acted to “alter[] Copilot to force it to provide mock PII.” (*Id.* ¶ 234.) But these allegations do not state a claim under the CCPA’s narrow private right of action. Nowhere in the complaint do Plaintiffs allege that any OpenAI Entity was “subject to an unauthorized access and exfiltration, theft, or disclosure” (b) “as a result of [their] violation of the duty to implement and maintain reasonable and appropriate security procedures and practices.” Cal. Civ. Code § 1798.150(a). Nor do plaintiffs provide any factual support for their allegation that any PII (much less their own PII) had been “exposed” or “reveal[ed].” (Comp. ¶¶ 233-34.) Plaintiffs’ CCPA claim has nothing to do with a data breach or unauthorized access to defendants’ network—the crux of what this statute was designed to protect.

#### 8. Plaintiffs Fail to State a Claim for Negligence.

Plaintiffs’ claim for negligent handling of personal data suffers from the same defects as their CCPA claim and must also be dismissed. To prevail on a negligence claim, a plaintiff must establish: “(1) defendant’s obligation to conform to a certain standard of conduct for the protection of others against unreasonable risks (duty); (2) failure to conform to the standard (breach of the duty); (3) a reasonably close connection between the defendant’s conduct and resulting injuries (proximate cause); and (4) actual loss (damages).” *Aguilar v. Hartford Accident & Indem. Co.*, No. CV 18-8123-R, 2019 WL 2912861, at \*2 (C.D. Cal. Mar. 13, 2019) (citing *Corales v. Bennett*, 567 F.3d 554, 572 (9th Cir. 2009)). The negligence pleading standard in the context of a data breach is “particularly demanding.” *See, e.g., In re Sony Gaming Networks and Customer Data Sec. Breach Litig.*, 996 F. Supp. 2d 942, 971-72 (S.D. Cal. 2014), *order corrected*, No. 11MD2258 AJB (MDD), 2014 WL 12603117 (S.D. Cal. Feb. 10, 2014) (“[T]he Court will not allow expensive, potentially burdensome class action discovery to ensue in the absence of a viable” negligence claim). Here, Plaintiffs have not pled any of the required elements.



First, the complaint fails to allege any duty of care the OpenAI Entities owed to Plaintiffs. Plaintiffs merely allege that the OpenAI Entities negligently “collect[ed], maintain[ed], and control[ed] their customers’ sensitive personal information.” (Compl. ¶ 237.) But Plaintiffs have not alleged that they are customers of any OpenAI Entity or had any relationship with an OpenAI Entity that would have created a duty. Nor have they alleged that the OpenAI Entities collected or maintained any of Plaintiffs’ “sensitive personal information” or “personal data.” (*Id.* ¶¶ 19-20, 23-28.) In fact, Plaintiffs allege that the data at issue is available in public Github repositories. (*Id.* ¶¶ 10, 46, 82-83.) Plaintiffs have failed to plead the existence of a duty owed by any OpenAI Entity. *Schmitt v. SN Servicing Corp.*, No. 21-CV-03355-WHO, 2021 WL 3493754, at \*4 (N.D. Cal. Aug. 9, 2021) (finding duty inadequately pled where plaintiffs fail to address the relevant factors, including the “foreseeability of harm to plaintiff, the degree of certainty that plaintiff suffered injury, the closeness of the connection between the defendant’s conduct and the injury suffered, ... the extent of the burden to defendant and the consequences to the community of imposing a duty with resulting liability for breach”) (cleaned up).

Second, Plaintiffs also have not shown a breach of any duty. Despite generic allegations that the OpenAI Entities learned of “multiple instances of release of [PII],” Plaintiffs offer no facts regarding these purported disclosures. (*Id.* ¶ 238.) On these threadbare allegations, Plaintiffs have not sufficiently alleged duty or breach. *See, e.g., Schmitt*, 2021 WL 3493754, at \*5 (dismissing negligence claim where plaintiffs failed to provide “factual allegations from which [the Court] can draw a reasonable inference that PII. . . was among the information compromised during the data breach); *Anderson*, 2019 WL 3753308, at \*5 (dismissing claim where “plaintiffs fail to allege any facts in support of their conclusory assertion” that defendant “fail[ed] to implement and maintain reasonable security procedures and practices”).

Plaintiffs’ citations to statutory standards—Cal. Civ. Code sections 1798.82, *et seq.*, and 1798.100, *et seq.*—and to the right to privacy under California Constitution—also fail to support a breach. Even if Plaintiffs can rely on these provisions to define the standard of care, Plaintiffs have not pled any facts demonstrating how OpenAI purportedly violated them. *See, e.g., Schmitt*, 2021 WL 3493754, at \*5; *Anderson*, 2019 WL 3753308, at \*5.

1 *Third*, Plaintiffs cannot establish proximate cause as they cannot make “the requisite  
2 connection between the alleged breach and damages.” *See Schmitt*, 2021 WL 3493754, at \*6  
3 (dismissing negligence claim because plaintiffs failed to show causation when “plaintiffs have not  
4 plausibly pleaded that PII or identifiable information was disclosed (information that [defendant]  
5 had a duty to protect)”).

6 *Fourth*, Plaintiffs fail to adequately plead any redressable injuries. None of Plaintiff’s  
7 alleged injuries satisfy the requisite damages standard for negligence. (*See* Compl. ¶ 239.)

8 • **Alleged loss of control over identity** is insufficient. *Aguilar*, 2019 WL 2912861 at \*2  
9 (“alleged loss of control [over] medical information and personal financial information” does not  
10 establish damages).

11 • **Alleged lost time** is “too speculative to constitute cognizable injury.” *Corona v. Sony*  
12 *Pictures Ent., Inc.*, No. 14-CV-09600 RGK (Ex), 2015 WL 3916744, at \*4 (C.D. Cal. June 15,  
13 2015). Plaintiffs fail to specify what steps they took “to cure [the] harm to their privacy.” (*See*  
14 Compl. ¶ 239.) In any event, “the cost of lost time” is “not recoverable under the economic loss  
15 doctrine.” *See Gardiner v. Walmart Inc.*, No. 20-CV-04618-JSW, 2021 WL 2520103, at \*8  
16 (N.D. Cal. Mar. 5, 2021) (time spent not recoverable in tort).

17 • **Alleged threat of future harm** “is insufficient to constitute actual loss.” *Corona*, 2015  
18 WL 3916744, at \*3; *see Huynh v. Quora, Inc.*, No. 18-CV-07597-BLF, 2020 WL 7408230, at \*6  
19 (N.D. Cal. June 1, 2020) (dismissing negligence claim because plaintiffs alleged the “mere danger  
20 of future harm,” without “specific factual statements that Plaintiffs’ [PII] has been misused”).

21 • **Alleged privacy injuries** are conclusory and inadequately pled. In any event, “an alleged  
22 loss of property in the form of personal information is insufficient to support a claim for  
23 negligence.” *Aguilar*, 2019 WL 2912861 at \*2.

24 • **Alleged economic loss** is unsupported by any facts demonstrating economic loss  
25 associated with the alleged disclosure of their PII. To the extent plaintiffs are attempting to allege  
26 economic losses, the economic loss rule bars recovery for any economic losses in tort unless there  
27 is a “special relationship” between the parties. *Corona*, 2015 WL 3916744, at \*5. Plaintiffs have  
28 not alleged that a special relationship exists here.



**9. Plaintiffs Fail to State a Civil Conspiracy Claim.**

Civil conspiracy “is not a cause of action, but a legal doctrine that imposes liability on persons who, although not actually committing a tort themselves, share with the immediate tortfeasors a common plan or design in its perpetration.” *Applied Equip. Corp. v. Litton Saudi Arabia Ltd.*, 7 Cal. 4th 503, 510-11 (1994). To plead a conspiracy, plaintiffs must allege “(1) the formation and operation of the conspiracy, (2) the wrongful act or acts done pursuant thereto, and (3) the damage resulting from such act or acts.” *Wasco Prods., Inc. v. Southwall Techs., Inc.*, 435 F.3d 989, 992 (9th Cir. 2006) (cleaned up). Further, such allegations “must be made within the sections of the complaint that contain plaintiff’s claims for the underlying torts.” *AccuImage Diagnostics Corp. v. Terarecon, Inc.*, 260 F. Supp. 2d 941, 947-48 (N.D. Cal. 2003)

Because a conspiracy charge seeks to hold parties liable for the conduct of others, courts impose a heightened standard at the pleading stage: “[P]laintiff must more clearly allege *specific action on the part of each defendant* that corresponds to the elements of a conspiracy cause of action” and “*cannot indiscriminately allege that conspiracies existed between and among all defendants.*” *AccuImage*, 260 F. Supp. 2d at 947-48 (emphasis added); *see also Ho Myung Moolsan Co. v. Manitou Mineral Water, Inc.*, 665 F. Supp. 2d 239, 256 (S.D.N.Y. 2009).

Plaintiffs’ conspiracy claim fails because this is not an independent cause of action. *See Accuimage*, 260 F. Supp. 2d at 947 (dismissing conspiracy claim with prejudice because “plaintiff cannot plead conspiracy as an independent cause of action”); *see also* 5 B. Witkin, Summary of California Law, Torts § 44 (9th ed.1988) (“Strictly speaking, however, there is no separate tort of civil conspiracy.”). Plaintiffs’ conspiracy claim should be dismissed on this ground alone.

Separately, the conspiracy claim fails because the complaint does not identify the role of each Defendant in the formation and operation of the alleged conspiracy, the wrongful acts done by each Defendant, or anything else about the alleged conspiracy, other than four paragraphs of boilerplate allegations. (*See* Compl. ¶¶ 241-244.) Plaintiffs allege vaguely that “[Defendants] agreed to a common plan or design to create, sell, and run Copilot to commit and conceal [various] tortious acts.” (*Id.* ¶ 244.) These allegations, which indiscriminately allege that conspiracies existed among *all Defendants*, do not satisfy the substantial showing required at the



1 pleading stage. *See e.g., AccuImage*, 260 F. Supp. at 947-48 (dismissing conspiracy claim with  
2 prejudice, in part, because plaintiff provided “only vague details” about the conspiracies).

3 In addition, the complaint does not satisfy the “wrongful acts” element upon which the  
4 conspiracy claim is based. To establish a “wrongful act,” plaintiffs must satisfy all elements of  
5 some other tort or wrong. *Gen. Am. Life Ins. Co. v. Rana*, 769 F. Supp. 1121, 1125 (N.D. Cal.  
6 1991). As detailed above, Plaintiffs cannot allege a violation of any of the laws on which they  
7 premise their civil conspiracy claim.

8 Finally, “agents and employees of a corporation cannot conspire with their corporate  
9 principal or employer where they act in their official capacities,” *AccuImage*, 260 F. Supp. 2d at  
10 947, nor can a corporate parent and subsidiary conspire together, *Laxalt v. McClatchy*, 622 F.  
11 Supp. 737, 745-46 (D. Nev. 1985) (dismissing conspiracy claim against corporation and its  
12 wholly owned subsidiaries and employees). Plaintiffs’ conspiracy claim against OpenAI, Inc. and  
13 its subsidiaries should be dismissed for this independent reason.

14 **10. Plaintiffs Fail to State a Claim for Declaratory Relief.**

15 Plaintiffs’ claim for declaratory relief should be dismissed because it is not an independent  
16 cause of action. *See Mayen v. Bank of Am. N.A.*, No. 14-CV-03757-JST, 2015 WL 179541, at \*5  
17 (N.D. Cal. Jan. 14, 2015). Where, as here, there is no basis for any of the underlying claims,  
18 dismissal is appropriate. *See Malasky v. Esposito*, No. 16-CV-04102-DMR, 2019 WL 79032, at  
19 \*10 (N.D. Cal. Jan. 2, 2019), *aff’d*, 781 F. App’x 643 (9th Cir. 2019) (dismissing declaratory relief  
20 claim “[b]ecause the court has dismissed [the] underlying claims”).

21 **V. CONCLUSION**

22 For all of these reasons, Plaintiffs’ complaint fails to state any claim against the OpenAI  
23 Entities. The complaint should be dismissed in its entirety.

1 Dated: January 26, 2023

MORRISON & FOERSTER LLP

2  
3 By: /s/ Michael A. Jacobs  
4 Michael A. Jacobs

5 MICHAEL A. JACOBS  
6 MJacobs@mofo.com  
7 JOSEPH C. GRATZ  
8 JGratz@mofo.com  
9 TIFFANY CHEUNG  
10 TCheung@mofo.com  
11 MORRISON & FOERSTER LLP  
12 425 Market Street  
13 San Francisco, California 94105-2482  
14 Telephone: (415) 268-7000  
15 Facsimile: (415) 268-7522

16 ROSE S. LEE  
17 RoseLee@mofo.com  
18 MORRISON & FOERSTER LLP  
19 707 Wilshire Boulevard  
20 Los Angeles, California 90017-3543  
21 Telephone: (213) 892-5200  
22 Facsimile: (213) 892-5454

23 Attorneys for Defendants OPENAI, INC., a  
24 Delaware nonprofit corporation, OPENAI,  
25 L.P., a Delaware limited partnership,  
26 OPENAI GP, L.L.C., a Delaware limited  
27 liability company, OPENAI STARTUP  
28 FUND GP I, L.L.C., a Delaware limited  
liability company, OPENAI STARTUP  
FUND I, L.P., a Delaware limited  
partnership, OPENAI STARTUP FUND  
MANAGEMENT, LLC, a Delaware limited  
liability company

# **EXHIBIT B**



Joseph R. Saveri (State Bar No. 130064)  
Steven N. Williams (State Bar No. 175489)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Louis A. Kessler (State Bar No. 243703)  
Elissa A. Buchanan (State Bar No. 249996)  
Travis Manfredi (State Bar No. 281779)  
**JOSEPH SAVERI LAW FIRM, LLP**  
601 California Street, Suite 1000  
San Francisco, California 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
Email: jsaveri@saverilawfirm.com  
swilliams@saverilawfirm.com  
czirpoli@saverilawfirm.com  
cyoung@saverilawfirm.com  
lkessler@saverilawfirm.com  
eabuchanan@saverilawfirm.com  
tmanfredi@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)  
1920 Hillhurst Avenue, #406  
Los Angeles, CA 90027  
Telephone: (323) 968-2632  
Facsimile: (415) 395-9940  
Email: mb@buttericklaw.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
OAKLAND DIVISION**

J. DOE 1, J. DOE 2, J. DOE 3, J. DOE 4, and J. DOE 5,  
individually and on behalf of all others similarly situated,

Individual and Representative Plaintiffs,

v.

GITHUB, INC., a Delaware corporation;  
MICROSOFT CORPORATION, a Washington  
corporation; OPENAI, INC., a Delaware nonprofit  
corporation; OPENAI, L.P., a Delaware limited  
partnership; OPENAI OPCO, L.L.C., a Delaware  
limited liability company; OPENAI GP, L.L.C., a  
Delaware limited liability company; OPENAI  
STARTUP FUND GP I, L.L.C., a Delaware limited

Case No. 4:22-cv-06823-JST  
4:22-cv-07074-JST

**FIRST AMENDED COMPLAINT**

**CLASS ACTION**

**DEMAND FOR JURY TRIAL**

4:22-cv-06823-JST

1 liability company; OPENAI STARTUP FUND I, L.P., a  
2 Delaware limited partnership; OPENAI STARTUP  
3 FUND MANAGEMENT, LLC, a Delaware limited  
4 liability company,

Defendants.

## TABLE OF CONTENTS

I.	OVERVIEW: A BRAVE NEW WORLD OF SOFTWARE PIRACY .....	1
II.	JURISDICTION AND VENUE.....	4
III.	INTRADISTRICT ASSIGNMENT .....	4
IV.	PARTIES.....	4
A.	Plaintiffs .....	4
B.	Defendants .....	6
V.	AGENTS AND CO-CONSPIRATORS .....	8
VI.	CLASS ALLEGATIONS.....	9
A.	Class Definitions.....	9
B.	Numerosity.....	10
C.	Typicality.....	10
D.	Commonality & Predominance.....	10
1.	DMCA Violations.....	10
2.	Contract-Related Conduct .....	11
3.	Unlawful-Competition Conduct .....	11
4.	Injunctive Relief.....	11
5.	Defenses .....	11
E.	Adequacy .....	12
F.	Other Class Considerations .....	12
VII.	FACTUAL ALLEGATIONS.....	12
A.	Introduction.....	12
B.	Codex Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses .....	13
C.	Copilot Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses .....	17
D.	Codex and Copilot Were Trained on Copyrighted Materials Offered Under Licenses.....	20
E.	Copilot Was Launched Despite Its Propensity for Producing Unlawful Outputs .....	21
F.	Copilot Reproduces the Code of the Named Plaintiffs Without Attribution.....	24



1.	Example: Copilot Outputs the Code of Doe 2 Essentially Verbatim .....	24
2.	Example: Copilot Outputs the Code of Doe 1 in Modified Format .....	26
3.	Example: Copilot Outputs the Code of Doe 5 In Modified Format .....	30
4.	Example: Copilot Outputs Code of Doe 5 Essentially Verbatim .....	33
G.	Codex and Copilot Were Designed to Withhold Attribution, Copyright Notices, and License Terms from Their Users .....	37
H.	Open-Source Licenses Began to Appear in the Early 1990s .....	39
I.	Microsoft Has a History of Flouting Open-Source License Requirements.....	41
J.	GitHub Was Designed to Cater to Open-Source Projects .....	43
K.	OpenAI Is Intertwined with Microsoft and GitHub.....	45
L.	Conclusion of Factual Allegations .....	47
VIII.	CLAIMS FOR RELIEF.....	48
IX.	DEMAND FOR JUDGMENT.....	64
X.	JURY TRIAL DEMANDED .....	66

1 Plaintiffs J. Doe 1, J. Doe 2, J. Doe 3, J. Doe 4 and J. Doe 5 (“Plaintiffs”), on behalf of  
2 themselves and all others similarly situated, bring this Class Action Complaint (the “Complaint”) against Defendants GitHub, Inc.; Microsoft Corporation; OpenAI, Inc.; OpenAI, L.P.; OpenAI  
3 OpCo, L.L.C.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI Startup Fund I,  
4 L.P.; and OpenAI Startup Fund Management, LLC<sup>1</sup> for violation of the Digital Millennium  
5 Copyright Act, 17 U.S.C. §§ 1201–1205 (the “DMCA”); breach of contract regarding the  
6 Suggested Licenses, breach of contract regarding GitHub’s policies including its terms of service;  
7 tortious interference with prospective economic relations; California’s Unfair Competition law,  
8 Cal. Bus. & Prof. Code section 17200, *et seq.*; common law unfair competition; negligence, and  
9 unjust enrichment.

## 11 I. OVERVIEW: A BRAVE NEW WORLD 12 OF SOFTWARE PIRACY

13 1. Plaintiffs and Class members are owners of copyright interests in materials made  
14 available publicly on GitHub that are subject to various licenses containing conditions for use of  
15 those works (the “Licensed Materials”). All the licenses at issue here (the “Licenses”) contain  
16 certain common terms (the “License Terms”).

17 2. “Artificial Intelligence” is referred to herein as “AI.” AI is defined for the  
18 purposes of this Complaint as a computer program that algorithmically simulates human  
19 reasoning or inference, often using statistical methods. Machine Learning (“ML”) is a subset of  
20 AI in which the behavior of the program is derived from studying a corpus of material called  
21 training data.

22 3. GitHub is a company founded in 2008 by a team of open-source enthusiasts. At  
23 the time, GitHub’s stated goal was to support open-source development, especially by hosting

---

24 <sup>1</sup> GitHub, Inc. is referred to as “GitHub.” Microsoft Corporation is referred to as “Microsoft.”  
25 OpenAI, Inc.; OpenAI, L.P.; OpenAI OpCo, L.L.C.; OpenAI GP, L.L.C.; OpenAI Startup Fund  
26 GP I, L.L.C.; OpenAI Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC are  
27 referred to collectively herein as “OpenAI.” Collectively, GitHub, Inc., Microsoft Corporation,  
28 OpenAI, Inc.; OpenAI, L.P.; OpenAI GP, L.L.C.; OpenAI Startup Fund GP I, L.L.C.; OpenAI  
Startup Fund I, L.P.; and OpenAI Startup Fund Management, LLC are referred to herein as  
“Defendants.”

1 open-source source code on the website github.com. Over the next 10 years, GitHub, based on  
2 these representations succeeded wildly, attracting nearly 25 million developers.

3 4. Developers published Licensed Materials on GitHub pursuant to written Licenses.  
4 In particular, the most popular ones share a common term: use of the Licensed Materials requires  
5 some form of *attribution*, usually by, among other things, including a copy of the license along  
6 with the name and copyright notice of the original author.

7 5. On October 26, 2018, Microsoft acquired GitHub for \$7.5 billion. Though some  
8 members of the open-source community were skeptical of this union, Microsoft repeated one  
9 mantra throughout: “Microsoft Loves Open Source.” For the first few years, Microsoft’s  
10 representations seemed credible.

11 6. Microsoft invested \$1 billion in OpenAI LP in July 2019 at a \$20 billion valuation.  
12 In 2020, Microsoft became exclusive licensee of OpenAI’s GPT-3 language model—despite  
13 OpenAI’s continued claims its products are meant to benefit “humanity” at large. In 2021,  
14 Microsoft began offering GPT-3 through its Azure cloud-computing platform. On October 20,  
15 2022, it was reported that OpenAI “is in advanced talks to raise more funding from Microsoft” at  
16 that same \$20 billion valuation. Copilot runs on Microsoft’s Azure platform. Microsoft has used  
17 Copilot to promote Azure’s processing power, particularly regarding AI.

18 7. On information and belief, Microsoft obtained a partial ownership interest in  
19 OpenAI in exchange for its \$1 billion investment. As OpenAI’s largest investor and largest service  
20 provider—specifically in connection with Microsoft’s Azure product—Microsoft exerts  
21 considerable control over OpenAI.

22 8. In June 2021, GitHub and OpenAI launched Copilot, an AI-based product that  
23 promises to assist software coders by providing or filling in blocks of code using AI. GitHub  
24 charges Copilot users \$10 per month or \$100 per year for this service. Copilot ignores, violates,  
25 and removes the Licenses offered by thousands—possibly millions—of software developers,  
26 thereby accomplishing software piracy on an unprecedented scale. Copilot outputs text derived  
27 from Plaintiffs’ and the Class’s Licensed Materials without adhering to the applicable License  
28 Terms and applicable laws. Copilot’s output is referred herein as “Output.”



1           9.       On August 10, 2021, OpenAI debuted its Codex product, which converts natural  
2 language into code and is integrated into Copilot. Copilot and Codex can be called either AIs or  
3 MLs. Codex and Copilot will be referred to as Ais herein unless a distinction is required.

4           10.      Though Defendants have been cagey about what data was used to train the AI,<sup>2</sup>  
5 they have conceded that the training data includes data in vast numbers of publicly accessible  
6 repositories on GitHub,<sup>3</sup> which include and are limited by Licenses.

7           11.      Among other things, Defendants stripped Plaintiffs' and the Class's attribution,  
8 copyright notice, and license terms from their code in violation of the Licenses and Plaintiffs' and  
9 the Class's rights. Defendants used Copilot to distribute the now-anonymized code to Copilot  
10 users as if it were created by Copilot.

11          12.      Copilot is run entirely on Microsoft's Azure cloud-computing platform.

12          13.      Copilot often simply reproduces code that can be traced back to open-source  
13 repositories or open-source licensees. Contrary to and in violation of the Licenses, code  
14 reproduced by Copilot *never* includes attributions to the underlying authors.

15          14.      GitHub and OpenAI have offered shifting accounts of the source and amount of  
16 the code or other data used to train and operate Copilot. They have also offered shifting  
17 justifications for why a commercial AI product like Copilot should be exempt from these license  
18 requirements, often citing "fair use."

19          15.      It is not fair, permitted, or justified. On the contrary, Copilot's goal is to replace a  
20 huge swath of open source by taking it and keeping it inside a GitHub-controlled paywall. It  
21 violates the licenses that open-source programmers chose and monetizes their code despite  
22 GitHub's pledge never to do so.

---

23  
24  
25          <sup>2</sup> "Training" an AI, as described in greater detail below, means feeding it large amounts of data  
26 that it interprets using given criteria. Feedback is then given to it to fine-tune its Output until it  
27 can provide Output with minimal errors.

28          <sup>3</sup> Repositories are containers for individual coding projects. They are where GitHub users upload  
their code and where other users can find it. Most GitHub users have multiple repositories.



## II. JURISDICTION AND VENUE

16. Plaintiffs bring this action on their own behalf as well as representatives of a Class of similarly situated individuals and entities. They seek to recover injunctive relief and damages as a result and consequence of Defendants' unlawful conduct.

17. Jurisdiction and venue are proper in this judicial district under 28 U.S.C. § 1331 pursuant to Defendants' violation of Section 1202(b) of the Digital Millennium Copyright Act, 17 U.S.C. §§ 1201–1205; and because a substantial part of the events giving rise to Plaintiffs' claims occurred in this District, a substantial portion of the affected interstate trade and commerce was carried out in this District, and three or more of the Defendants reside in this District and/or are licensed to do business in this District. Each Defendant has transacted business, maintained substantial contacts, and/or committed overt acts in furtherance of the illegal scheme and conspiracy throughout the United States, including in this District. Defendants' conduct has had the intended and foreseeable effect of causing injury to persons residing in, located in, or doing business throughout the United States, including in this District.

## III. INTRADISTRICT ASSIGNMENT

18. Pursuant to Civil Local Rule 3.2 (c) and (e), assignment of this case to the San Francisco Division of the United States District Court for the Northern District of California is proper because a substantial amount of the development of the Copilot product as well as of the interstate trade and commerce involved and affected by Defendants' conduct giving rise to the claims herein occurred in this Division. Furthermore, Defendants GitHub and all the OpenAI entities are headquartered within this Division.

## IV. PARTIES

### A. PLAINTIFFS

19. Plaintiff J. Doe 1, [REDACTED], is a resident of the State of New Hampshire. Plaintiff Doe 1 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 1 has published Licensed Materials they claim a copyright interest in under the following Suggested Licenses:

MIT License and GNU General Public License version 3.0. Plaintiff was, and continues to be, injured during the Class Period as a result of Defendants' unlawful conduct alleged herein.

20. Plaintiff J. Doe 2, [REDACTED], is a resident of the State of Illinois. Plaintiff Doe 2 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 2 has published Licensed Materials they claim a copyright interest in under the following Suggested Licenses: MIT License; GNU General Public License version 3.0; GNU Affero General Public License version 3.0; The 3-Clause BSD License; and Apache License 2.0. Plaintiff was, and continues to be, injured during the Class Period as a result of Defendants' unlawful conduct alleged herein.

21. Plaintiff J. Doe 3, [REDACTED], is a resident of the State of Idaho. Plaintiff Doe 3 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 3 has published Licensed Materials they claim a copyright interest in under the following Suggested Licenses: MIT License; GNU General Public License version 3.0; and GNU Affero General Public License version 3.0. Plaintiff was, and continues to be, injured during the Class Period as a result of Defendants' unlawful conduct alleged herein.

22. Plaintiff J. Doe 4, [REDACTED], is a resident of the State of South Carolina. Plaintiff Doe 4 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 4 has published Licensed Materials they claim a copyright interest in under the following Suggested Licenses: GNU General Public License v2.0 and GNU General Public License v3.0. Plaintiff was, and continues to be, injured during the Class Period as a result of Defendants' unlawful conduct alleged herein.

23. Plaintiff J. Doe 5, [REDACTED], is a resident of the Commonwealth of Massachusetts. Plaintiff Doe 5 published Licensed Materials they owned a copyright interest in to at least one GitHub repository under one of the Suggested Licenses. Specifically, Doe 5 has published Licensed Materials they claim a copyright interest in under the following Suggested Licenses: MIT License; Apache License 2.0; and GNU General Public License v3.0.

**B. DEFENDANTS**

24. Defendant GitHub, Inc. is a Delaware corporation with its principal place of business located at 88 Colin P Kelly Jr Street, San Francisco, CA 94107. GitHub sells, markets, and distributes Copilot throughout the internet and other sales channels throughout the United States, including in this District. GitHub released Copilot on a limited “technical preview” basis on June 29, 2021. On June 21, 2022, Copilot was released to the public as a subscription-based service for individual developers. GitHub is a party to the unlawful conduct alleged herein.

25. Defendant Microsoft Corporation is a Washington corporation with its principal place of business located at One Microsoft Way, Redmond, Washington 98052. Microsoft announced its acquisition of Defendant GitHub, Inc. on June 4, 2018. On October 26, 2018, Microsoft finalized its acquisition of GitHub. Microsoft owns and operates GitHub. Through its corporate ownership, control of the GitHub Board of Directors, active management, and other means, Microsoft sells, markets, and distributes Copilot. Microsoft is a party to the unlawful conduct alleged herein.

26. Defendant OpenAI, Inc. is a Delaware nonprofit corporation with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI, Inc. is a party to the unlawful conduct alleged herein. It—along with OpenAI, L.P.—programed, trained, and maintains Codex, which infringes all the same rights at Copilot and is also an integral piece of Copilot. Copilot requires Codex to function. OpenAI, Inc. is a party to the unlawful conduct alleged herein. OpenAI, Inc. founded, owns, and exercises control over all the other OpenAI entities, including those set forth in Paragraphs 27–32.

27. Defendant OpenAI, L.P. is a Delaware limited partnership with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI, L.P. is a party to the unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI, L.P. is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI, L.P. is the OpenAI entity that co-created Copilot and offers it jointly with GitHub. OpenAI’s revenue, including revenue from Copilot, is received by OpenAI, L.P. OpenAI, Inc. controls OpenAI, L.P. directly and through the other OpenAI entities.



28. Defendant OpenAI OpCo, L.L.C. is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI OpCo, L.L.C. is a party to the unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI OpCo, L.L.C. is a wholly owned subsidiary of OpenAI, Inc. that is operated for profit. OpenAI OpCo, L.L.C. is the OpenAI entity that co-created Copilot and offers it jointly with GitHub. OpenAI's revenue, including revenue from Copilot, is received by OpenAI OpCo, L.L.C. OpenAI, Inc. controls OpenAI OpCo, L.L.C. directly and through the other OpenAI entities.

29. Defendant OpenAI GP, L.L.C. ("OpenAI GP") is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI GP is the general partner of OpenAI, L.P. OpenAI GP manages and operates the day-to-day business and affairs of OpenAI, L.P. OpenAI GP is liable for the debts, liabilities and obligations of OpenAI, L.P., including litigation and judgments. OpenAI GP is a party to the unlawful conduct alleged herein. Its primary activity is research and technology. OpenAI GP is the general partner of OpenAI, L.P. OpenAI GP was aware of the unlawful conduct alleged herein *and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI, Inc. directly controls* OpenAI GP.

30. Defendant OpenAI Startup Fund I, L.P. ("OpenAI Startup Fund I") is a Delaware limited partnership with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund I was instrumental in the foundation of OpenAI, L.P., including the creation of its business strategy and providing initial funding. Through participation in OpenAI Startup Fund I, certain entities and individuals obtained an ownership interest in OpenAI, L.P. Plaintiffs are informed and believed, and on that basis allege that OpenAI Startup Fund I participated in the organization and operation of OpenAI, L.P. OpenAI Startup Fund I is a party to the unlawful conduct alleged herein. OpenAI Startup Fund I was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

31. Defendant OpenAI Startup Fund GP I, L.L.C. ("OpenAI Startup Fund GP I") is a Delaware limited liability company with its principal place of business located at 3180 18th

Street, San Francisco, CA 94110. OpenAI Startup Fund GP I is the general partner of OpenAI Startup Fund I. OpenAI Startup Fund GP I manages and operates the day-to-day business and affairs of OpenAI Startup Fund I. OpenAI Startup Fund GP I is liable for the debts, liabilities and obligations of OpenAI Startup Fund I, including litigation and judgments. OpenAI Startup Fund GP I was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period. OpenAI Startup Fund GP I is a party to the unlawful conduct alleged herein. Sam Altman, co-founder, CEO, and Board member of OpenAI, Inc. is the Manager of OpenAI Startup Fund GP I. OpenAI Startup Fund GP I is the General Partner of OpenAI Startup Fund I, L.P.

32. Defendant OpenAI Startup Fund Management, LLC (“OpenAI Startup Fund Management”) is a Delaware limited liability company with its principal place of business located at 3180 18th Street, San Francisco, CA 94110. OpenAI Startup Fund Management is a party to the unlawful conduct alleged herein. OpenAI Startup Fund Management was aware of the unlawful conduct alleged herein and exercised control over OpenAI, L.P. throughout the Class Period.

## V. AGENTS AND CO-CONSPIRATORS

33. The unlawful acts alleged against the Defendants in this class action complaint were authorized, ordered, or performed by the Defendants’ respective officers, agents, employees, representatives, or shareholders while actively engaged in the management, direction, or control of the Defendants’ businesses or affairs.

34. The Defendants’ agents operated under the explicit and apparent authority of their principals.

35. Each Defendant, and its subsidiaries, affiliates and agents operated as a single unified entity.

36. Various persons and/or firms not named as Defendants herein may have participated as coconspirators in the violations alleged herein and may have performed acts and made statements in furtherance thereof.

37. Each acted as the principal, agent, or joint venture of, or for other Defendants with respect to the acts, violations, and common course of conduct alleged herein.

## VI. CLASS ALLEGATIONS

### A. Class Definitions

38. Plaintiffs bring this action for damages and injunctive relief on behalf of themselves and all others similarly situated as a class action pursuant to Rules 23(a), 23(b)(2), and 23(b)(3) of the Federal Rules of Civil Procedure, on behalf of the following Classes:

#### **“Injunctive Relief Class” under Rule 23(b)(2):**

All persons or entities domiciled in the United States that, (1) owned an interest in at least one US copyright in any work; (2) offered that work under one of GitHub’s Suggested Licenses<sup>4</sup>; and (3) stored Licensed Materials in any public GitHub repositories at any time between January 1, 2015 and the present (the “Class Period”).

#### **“Damages Class” under Rule 23(b)(3):**

All persons or entities domiciled in the United States that, (1) owned an interest in at least one US copyright in any work; (2) offered that work under one of GitHub’s Suggested Licenses; and (3) stored Licensed Materials in any public GitHub repositories at any time during the Class Period.

These “Class Definitions” specifically exclude the following person or entities:

---

<sup>4</sup> When a GitHub user creates a new repository, they have the option of selecting one of thirteen licenses from a dropdown menu to apply to the contents of that repository. (They can also apply a different license later, or no license.) The Creative Commons Zero v1.0 Universal and the Unlicense donate the covered work to the public domain and/or otherwise waive all copyrights and related rights. Because they do not contain the necessary provisions nor do they even allow the owner to make copyright claims in most circumstances, they are not included in the Class Definition. We refer to the remaining eleven options as the “Suggested Licenses,” which are: (1) Apache License 2.0 (“Apache 2.0”); (2) GNU General Public License version 3 (“GPL-3.0”); (3) MIT License (“MIT”); (4) The 2-Clause BSD License (“BSD 2”); (5) The 3-Clause BSD License (“BSD 3”); (6) Boost Software License (“BSL-1.0”); (7) Eclipse Public License 2.0 (“EPL-2.0”); (8) GNU Affero General Public License version 3 (“AGPL-3.0”); (9) GNU General Public License version 2 (“GPL-2.0”); (10) GNU Lesser General Public License version 2.1 (“LGPL-2.1”); and (11) Mozilla Public License 2.0 (“MPL-2.0”). These Suggested Licenses each contain at least three common requirements for use of the Licensed Materials in a derivative work or copy: attribution to the owner of the Licensed Materials (“Attribution”), inclusion of a copyright notice (“Copyright Notice”), and inclusion of the applicable Suggested License’s text (“License Terms”).



- a. Any of the Defendants named herein;
- b. Any of the Defendants' co-conspirators;
- c. Any of Defendants' parent companies, subsidiaries, and affiliates;
- d. Any of Defendants' officers, directors, management, employees, subsidiaries, affiliates, or agents;
- e. All governmental entities; and
- f. The judges and chambers staff in this case, as well as any members of their immediate families.

**B. Numerosity**

39. Plaintiffs do not know the exact number of Class members, because such information is in the exclusive control of Defendants. Plaintiffs are informed and believe that there are at least thousands of Class members geographically dispersed throughout the United States such that joinder of all Class members in the prosecution of this action is impracticable.

**C. Typicality**

40. Plaintiffs' claims are typical of the claims of their fellow Class members because Plaintiffs and Class members all own code published under a License. Plaintiffs and the Class published work subject to a License to GitHub later used by Copilot. Plaintiffs and absent Class members were damaged by this and other wrongful conduct of Defendants as alleged herein. Damages and the other relief sought herein is common to all members of the Class.

**D. Commonality & Predominance**

41. Numerous questions of law or fact common to the entire Class arise from Defendants' conduct—including, but not limited to those identified below:

**1. DMCA Violations**

- Whether Defendants' conduct violated the Class's rights under the DMCA when GitHub and OpenAI caused Codex and Copilot to ingest and distribute Licensed Materials without including any associated Attribution, Copyright Notice, or License Terms.

1           **2. Contract-Related Conduct**

- 2           • Whether Defendants violated the Licenses governing use of the Licensed
- 3           Materials by using them to train Copilot and for republishing those materials
- 4           without appending the required Attribution, Copyright Notice, or License
- 5           Terms.
- 6           • Whether Defendants interfered in prospective economic relations between the
- 7           Class and the public regarding the Licensed Materials by concealing the
- 8           License Terms.
- 9           • Whether Defendants intentionally or negligently interfered with a prospective
- 10          economic advantage.

11           **3. Unlawful-Competition Conduct**

- 12          • Whether Defendants passed-off the Licensed Materials as its own creation
- 13          and/or Codex or Copilot's creation.
- 14          • Whether Defendants were unjustly enriched by the unlawful conduct alleged
- 15          herein.
- 16          • Whether Defendants' conduct alleged herein constitutes Unfair Competition
- 17          under California Business and Professions Code section 17200 *et seq.*
- 18          • Whether Defendants' conduct alleged herein constitutes unfair competition
- 19          under the common law.

20           **4. Injunctive Relief**

- 21          • Whether this Court should enjoin Defendants from engaging in the unlawful
- 22          conduct alleged herein. And what the scope of that injunction would be.

23           **5. Defenses**

- 24          • Whether any affirmative defense excuses Defendants' conduct.
- 25          • Whether any statutes of limitation limit Plaintiffs' and the Class's potential for
- 26          recovery.
- 27          • Whether any applicable statutes of limitation should be tolled as a result of
- 28          Defendants' fraudulent concealment of their unlawful conduct.

42. These and other questions of law and fact are common to the Class and predominate over any questions affecting the Class members individually.

**E. Adequacy**

43. Plaintiffs will fairly and adequately represent the interests of the Class because they have experienced the same harms as the Class and have no conflicts with any other members of the Class. Furthermore, Plaintiffs have retained sophisticated and competent counsel (“Class Counsel”) who are experienced in prosecuting Federal and state class actions throughout the United States and other complex litigation and have extensive experience advising clients and litigating intellectual property, competition, contract, and privacy matters.

**F. Other Class Considerations**

44. Defendants have acted on grounds generally applicable to the Class, thereby making final injunctive relief appropriate with respect to the Class as a whole.

45. This class action is superior to alternatives, if any, for the fair and efficient adjudication of this controversy. Prosecuting the claims pleaded herein as a class action will eliminate the possibility of repetitive litigation. There will be no material difficulty in the management of this action as a class action.

46. The prosecution of separate actions by individual Class members would create the risk of inconsistent or varying adjudications, establishing incompatible standards of conduct for Defendants.

**VII. FACTUAL ALLEGATIONS**

**A. Introduction**

47. This class action against Defendants concerns an OpenAI product called Codex and a GitHub product called Copilot.

48. OpenAI began development of Codex sometime after OpenAI was founded in December 2015 and released Codex on a limited basis in August 2021.

49. GitHub began development of Copilot sometime in 2019, released it on a limited basis in June 2021, and released it as a paid subscription service in June 2022.



50. Codex and Copilot are assistive AI-based systems offered to software programmers. These AI systems are each trained on a large corpus of publicly accessible software code and other materials, including all the Licensed Materials. Defendants describe these products as making coding easier by accepting a code “prompt” from a programmer and emitting a possible completion of that code, which is referred to herein as Output, as set forth above. Defendants claim Codex and Copilot do not retain copies of the materials they are trained on. In practice, however, the Output is often a near-identical reproduction of code from the training data.

51. Codex and Copilot are related. Copilot is a joint venture between GitHub and OpenAI. Codex is a standalone product released by OpenAI that also “powers GitHub Copilot, which [OpenAI] built and launched in partnership with GitHub.”<sup>5</sup> “GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor.”<sup>6</sup> The version of Codex used to power Copilot is modified from OpenAI’s standalone version. Defendants have kept secret the details of Codex’s modifications and its integration into or interaction with Copilot.

**B. Codex Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses**

52. Below is an explanation of how Codex functions. When Codex is prompted with:

```
function isEven(n) {
```

it assumes this is the beginning of a function written in the JavaScript language that will test whether a number is even.

---

<sup>5</sup> See <https://openai.com/blog/openai-codex/>.

<sup>6</sup> See <https://github.com/features/copilot/>.

53. Based on this assumption, Codex will then provide Output meant to complete the rest of the function. Based on the given prompt, it produced the following response:<sup>7</sup>

```
function isEven(n) {  
  if (n == 0)  
    return true;  
  else if (n == 1)  
    return false;  
  else if (n < 0)  
    return isEven(-n);  
  else  
    return isEven(n - 2);  
}  
console.log(isEven(50));  
// → true  
console.log(isEven(75));  
// → false  
console.log(isEven(-1));  
// → ??
```

54. The function itself occupies the first ten lines. Six additional lines follow the function, beginning with “`console.log(isEven(50))`”. One possible explanation for Codex’s inclusion of these lines is to test the “`isEven`” function. Though not part of the function itself, the lines will confirm the function works for certain values. In this case, the code implies that “`isEven(50)`” should return the value “`true`”, and “`isEven(75)`” should return “`false`”. Those answers are correct.

55. The penultimate line indicates “`isEven(-1)`” should return “`??`”. This is an error, as “`isEven(-1)`” should return “`false`”.

56. Codex cannot and does not understand the meaning of software code or any other Licensed Materials. But in training, what became Codex was exposed to an enormous amount of existing software code (its “Training Data”) and—with input from its trainers and its own internal processes—inferred certain statistical patterns governing the structure of code and other Licensed Materials. The finished version of Codex, once trained, is known as a “Model.”

---

<sup>7</sup> Due to the nature of Codex, Copilot, and AI in general, Plaintiffs cannot be certain these examples would produce the same results if attempted following additional trainings of Codex and/or Copilot. However, these examples are representative of Codex and Copilot’s Output at the time just prior to the filing of this Complaint.

57. When given a prompt, such as the initial prompt discussed above—“function  
isEven(n) {”—Codex identifies the most statistically likely completion, based on the  
examples it reviewed in training. Every instance of Output from Codex is derived from material in  
its Training Data. Most of its Training Data consisted of Licensed Materials.

58. Codex does not “write” code the way a human would, because it does not  
understand the meaning of code. Codex’s lack of understanding of code is evidenced when it  
emits extra code that is not relevant under the circumstances. Here, Codex was only prompted to  
produce a function called “isEven”. To produce its answer, Codex relied on Training Data that  
also appended the extra testing lines. Having encountered this function and the follow-up lines  
together frequently, Codex extrapolates they are all part of one function. A human with even a  
basic understanding of how JavaScript works would know the extra lines are not part of the  
function itself.

59. Beyond the superfluous and inaccurate extra lines, this “isEven” function also  
contains two major defects. First, it assumes the variable “n” holds an integer. It could contain  
some other kind of value, like a decimal number or text string, which would cause an error.  
Second, even if “n” does hold an integer, the function will trigger a memory error called a “stack  
overflow” for sufficiently large integers. For these reasons, experienced programmers would not  
use Codex’s Output.

60. Codex does not identify the owner of the copyright to this Output, nor any  
other—it has not been trained to provide Attribution. Nor does it include a Copyright Notice nor  
any License Terms attached to the Output. This is by design—Codex was not coded or trained to  
track or reproduce such data. The Output in the example above is taken from *Eloquent JavaScript*  
by Marijn Haverbeke.<sup>8</sup>

---

<sup>8</sup> <https://eloquentjavascript.net/code/#3.2>. *Eloquent JavaScript* is “Licensed under a Creative Commons [A]tribution-[N]oncommercial license. All code in this book may also be considered licensed under an MIT license.” See <https://eloquentjavascript.net/>. Thus, having also been posted on GitHub, the code Codex relied on meets the definition of Licensed Materials.



61. Here is the exercise from *Eloquent JavaScript*:

```
// Your code here.

console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → ??
```

62. The exercise includes the “??” error. However, for Haverbeke’s purposes, this is not an error but a placeholder value for the reader to fill in. Codex—as a mere probabilistic model—fails to recognize this nuance. The inclusion of the double question marks confirms unequivocally that Codex took this code directly from a copyrighted source without following any of the attendant License Terms.

63. Haverbeke provides the following solution to the function discussed above:

```
function isEven(n) {
  if (n == 0) return true;
  else if (n == 1) return false;
  else if (n < 0) return isEven(-n);
  else return isEven(n - 2);
}

console.log(isEven(50));
// → true
console.log(isEven(75));
// → false
console.log(isEven(-1));
// → false
```

64. Aside from different line breaks—which are not semantically meaningful in JavaScript—this code for the function “isEven” is the same as what Codex produced. The tests are also the same, though in this case Haverbeke provides the right answer for “isEven(-1)”, which is “false”. Codex has reproduced Haverbeke’s Licensed Material almost verbatim, with the only difference being drawn from a different portion of those same Licensed Materials.

65. There are many copies of Haverbeke’s code stored in public repositories on GitHub, where programmers who are working through Haverbeke’s book store their answers.

66. The MIT license provides that “The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.”<sup>9</sup> Any person taking this code directly from *Eloquent JavaScript* would have direct access to these License Terms and know to follow them if incorporating the Licensed Materials into a derivative work and/or copying them. Codex does not provide these License Terms.

67. OpenAI Codex’s Output would frequently, perhaps even constantly, contain Licensed Materials, i.e., it would have conditions associated with it through its associated license. In its 2021 research paper about Codex called “Evaluating Large Language Models Trained on Code,” OpenAI stated Codex’s Output is “often incorrect” and can contain security vulnerabilities and other “misalignments” (meaning, departures from what the user requested).

68. Most open-source licenses require attribution of the author, notice of their copyright, and a copy of the license specifically to ensure that future coders can easily credit all previous authors and ensure they adhere to all applicable licenses. All the Suggested Licenses include these requirements.

69. Ultimately, Codex derives its value primarily from its ability to locate and output potentially useful Licensed Materials. And from its obfuscation of any rights associated with those materials.

**C. Copilot Outputs Copyrighted Materials Without Following the Terms of the Applicable Licenses**

70. GitHub Copilot works in a similar way to OpenAI Codex. As mentioned above, a modified version of Codex is used as the engine that powers Copilot.

71. Copilot is installed by the end user as an extension to various code editors, including Microsoft’s Visual Studio and VS Code. As the user types into the editor, their code is uploaded in real time to Microsoft’s Azure cloud platform, where they become prompts for Copilot.

---

<sup>9</sup> See Appendix A for full text of the MIT License.

72. When we give Copilot the same prompt discussed above in Paragraph 52, “function isEven(n) {”, it interprets the prompt as the beginning of a function written in the JavaScript language that will test whether a number is even, same as Codex.

73. However, the Output of Copilot in response to the prompt is different than Codex, namely:

```
function isEven(n) {  
    return n % 2 === 0;  
}
```

74. This function is much closer to what a human programmer might use as compared to Codex’s older, inaccurate offering. It handles all values and types of “n” correctly. It does not cause a stack overflow for larger values of “n” like the Codex Output.

75. Copilot’s Output, like Codex’s, is derived from existing code. Namely, sample code that appears in the online book *Mastering JS*, written by Valeri Karpov.<sup>10</sup> Like Codex’s Output, Copilot’s is also based upon copyrighted educational material. *Mastering JS* is a set of educational exercises for programmers. Like *Eloquent JavaScript*, there are many copies of Karpov’s exercise stored in public repositories on GitHub. Programmers working through *Mastering JS* store their answers there.

76. If Copilot is prompted with the name of a function that will test whether a number is prime (that is, a number that can only be evenly divided by 1 and itself), namely “function isPrime(n) {”, it returns:

```
function isPrime(n) {  
    if (n < 2) {  
        return false;  
    }  
    for (let i = 2; i < n; i++) {  
        if (n % i === 0) {  
            return false;  
        }  
    }  
    return true;  
}
```

---

<sup>10</sup> <https://masteringjs.io/tutorials/fundamentals/modulus/>.



77. Though this function will work, it contains an error often made by beginner programmers that makes it much slower than it could be. Namely, the loop in the middle, which checks possible divisors, does not need to check every divisor smaller than “n,” only the divisors smaller than the square root of “n”. As with Codex, Copilot has no understanding of how the code works. It knows that more functions called “isPrime” contain the portion that checks for all divisors smaller than “n”, so that is what it offers. It does not return what it “thinks” is best, it returns what it has seen *the most*. It is not writing, it is reproducing (i.e., copying).

78. Like the other examples above—and most of Copilot’s Output—this output is nearly a verbatim copy of copyrighted code. In this case, it is substantially similar to the “isPrime” function in the book *Think JavaScript* by Matthew X. Curinga et al,<sup>11</sup> which is:

```
function isPrime(n) {  
  if (n < 2) {  
    return false;  
  }  
  for (let i = 2; i < n; i++) {  
    if (n % i === 0) {  
      return false;  
    }  
  }  
  return true;  
}
```

79. As with the other examples above, the source of Copilot’s Output is a programming textbook. Also like the books the other examples were taken from, there are many copies of Curinga’s code stored in public repositories on GitHub where programmers who are working through Curinga’s book keep copies of their answers.

80. The material in Curinga’s book is made available under the GNU Free Documentation License. Although this is not one of the Suggested Licenses, it contains similar attribution provisions, namely that “You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and

---

<sup>11</sup> <https://matt.curinga.com/think-js/#solving-problems-with-for-loops>.

the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License.”<sup>12</sup>

81. As with Codex, Copilot does not provide the end user any attribution of the original author of the code, nor anything about their license requirements. There is no way for the Copilot user to know that they must provide attribution, copyright notice, nor a copy of the license’s text. And with regard to the GNU Free Documentation License, Copilot users would not be aware that they are limited in what conditions they can place on the use of derivative works they make using this copyrighted code. Had the Copilot user found this code in a public GitHub repository or a copy of the book it was originally published in, they would find the GNU Free Documentation License at the same time and be aware of its terms. Copilot finds that code for the user but excises the license terms, copyright notice, and attribution. This practice allows its users to assume that the code can be used without restriction. It cannot.

**D. Codex and Copilot Were Trained on Copyrighted Materials Offered Under Licenses**

82. Codex is an AI system. Another way to describe it is a “model.” Without Codex, Copilot, or another AI-code-lookup-tool, code is written both by originating code from the writer’s own knowledge of how to write code as well as by finding pre-written portions of code that—under the terms of the applicable license—may be incorporated into the coding project.

83. Unlike a human programmer that has learned how code works and notices when code it is copying has attached license terms, a copyright notice, and/or attribution, Codex and Copilot were developed by feeding a corpus of material, called “training data,” into them. These AI programs ingest all the data and, through a complex probabilistic process, predict what the most likely solution to a given prompt a user would input is. Though more complicated in practice, essentially Copilot returns the solution it has found in the most projects when those projects are somehow weighted to adjust for whatever variables Codex or Copilot have identified as relevant.

---

<sup>12</sup> <https://matt.curinga.com/think-js/#gnu-free-documentation-license>.

84. Codex and Copilot were not programmed to treat attribution, copyright notices, and license terms as legally essential. Defendants made a deliberate choice to expedite the release of Copilot rather than ensure it would not provide unlawful Output.

85. The words “study” and “training” and “learning” in connection with AI describe algorithmic processes that are not analogous to human reasoning. AI models cannot “learn” as humans do, nor can it “understand” semantics and context the way humans do. Rather, it detects statistically significant patterns in its training data and provides Output derived from its training data when statistically appropriate. A “brute force” approach like this would not be efficient nor even possible for humans. A human could not memorize, statistically analyze, and easily access thousands of gigabytes of existing code, a task now possible for powerful computers like those that make up Microsoft’s Azure cloud platform. To accomplish the same task, a human may search for Licensed Materials that serve their purpose if they believe such materials exist. And if that human finds such materials, they will probably abide by its License Terms rather than risk infringing its owners’ rights. At the very least, if they incorporate those Licensed Materials into their own project without following its terms they will be doing so knowingly.

**E. Copilot Was Launched Despite Its Propensity for Producing Unlawful Outputs**

86. GitHub and OpenAI have not provided much detail regarding what data Codex and OpenAI were trained on. Plaintiffs know for certain from GitHub and OpenAI’s statements, that both systems were trained on publicly available GitHub repositories, with Copilot having been trained on all available public GitHub repositories.

87. According to OpenAI, Codex was trained on “billions of lines of source code from publicly available sources, including code in public GitHub repositories.” Similarly, GitHub has described<sup>13</sup> Copilot’s training material as “billions of lines of public code.” GitHub researcher Eddie Aftandilian confirmed in a recent podcast<sup>14</sup> that Copilot is “train[ed] on public repos on GitHub.”

---

<sup>13</sup> <https://github.blog/2021-06-30-github-copilot-research-recitation/>.

<sup>14</sup> <https://www.se-radio.net/2022/10/episode-533-eddie-aftandilian-on-github-copilot/>.



88. In a recent customer-support message, GitHub’s support department clarified certain facts about training Copilot. First, GitHub said that “training for Codex (the model used by Copilot) is done by OpenAI, not GitHub.” Second, in its support message, GitHub put forward a more detailed justification for its use of copyrighted code as training data:

Training machine learning models on publicly available data is considered fair use across the machine learning community . . . OpenAI’s training of Codex is done in accordance with global copyright laws which permit the use of publicly accessible materials for computational analysis and training of machine learning models, and do not require consent of the owner of such materials. Such laws are intended to benefit society by enabling machines to learn and understand using copyrighted works, much as humans have done throughout history, and to ensure public benefit, these rights cannot generally be restricted by owners who have chosen to make their materials publicly accessible.

The claim that training ML models on publicly available code is widely accepted as fair use is not true. And regardless of this concept’s level of acceptance in “the machine learning community,” under Federal law, it is illegal.

89. Former GitHub CEO Nat Friedman said in June 2021—when Copilot was released to a limited number of customers—that “training ML systems on public data is fair use.”<sup>15</sup> Friedman’s statement is pure speculation; no Court has considered the question of whether “training ML systems on public data is fair use.” The Fair Use affirmative defense is only applicable to Section 501 copyright infringement. It is not a defense to violations of the DMCA, breach of contract, nor any other claim alleged herein. It cannot be used to avoid liability here. At the same time Friedman asserted “the output [of Copilot] belongs to the operator.”

90. Other open-source stakeholders have made this point already. For example, in June 2021, Software Freedom Conservancy (“SFC”), a prominent open-source advocacy organization, asked Microsoft and GitHub to provide “legal references for GitHub’s public legal positions.” No references were provided by any of the Defendants.<sup>16</sup>

---

<sup>15</sup> <https://twitter.com/natfriedman/status/1409914420579344385/>.

<sup>16</sup> <https://sfconservancy.org/blog/2022/feb/03/github-copilot-copyleft-gpl/>.

91. Beyond the examples above, Copilot regularly Output’s verbatim copies of Licensed Materials. For example, Copilot reproduced verbatim well-known code from the game Quake III, use of which is governed by one of the Suggested Licenses—GPL-2.<sup>17</sup>

92. Copilot also reproduced code that had been released under a license that allowed its use only for free games and required attribution by including a copy of the license. Copilot did not mention nor include the underlying license when providing a copy of this code as Output.<sup>18</sup>

93. Texas A&M computer-science professor Tim Davis has provided numerous examples of Copilot reproducing code belonging to him without its license or attribution.<sup>19</sup>

94. GitHub concedes that in ordinary use, Copilot will reproduce passages of code verbatim: “Our latest internal research shows that about 1% of the time, a suggestion [Output] may contain some code snippets longer than ~150 characters that matches” code from the training data. This standard is more limited than is necessary for copyright infringement. But even using GitHub’s own metric and the most conservative possible criteria, Copilot has violated the DMCA at least tens of thousands of times.

95. In June 2022, Copilot had 1,200,000 users. If only 1% of users have ever received Output based on Licensed Materials and only once each, Defendants have “only” breached Plaintiffs’ and the Class’s Licenses 12,000 times. However, each time Copilot outputs Licensed Materials without attribution, the copyright notice, or the License Terms it violates the DMCA three times. Thus, even using this extreme underestimate, Copilot has “only” violated the DMCA 36,000 times.<sup>20</sup> Because Copilot constantly Outputs code as a user writes, and because nearly all of Copilot’s training data was Licensed Material, this number is most likely exponentially lower than the true number of breaches and DMCA violations.

---

<sup>17</sup> <https://twitter.com/stefankarpinski/status/1410971061181681674/>.

<sup>18</sup> <https://twitter.com/ChrisGr93091552/status/1539731632931803137/>.

<sup>19</sup> <https://twitter.com/DocSparse/status/1581461734665367554/>.

<sup>20</sup> These violations of Section 1202 of the DMCA each incur statutory damages of “not less than \$2,500 or more than \$25,000.” 17 U.S.C. § 1203(c)(3)(B). This extremely conservative estimate of Defendants’ number of direct violations translates to \$90 million to \$900 million in statutory damages.

96. Furthermore, the Suggested Licenses impose attribution obligations not only when Licensed Materials have been used verbatim, but also when Licensed Materials have been modified or adapted. Though Output from Copilot is often a verbatim copy, even more often it is a modification: for instance, a near-identical copy that contains only semantically insignificant variations of the original Licensed Materials, or a modified copy that recreates the same algorithm. Whenever Copilot outputs Licensed Materials in a manner that qualifies as a modification, the attribution requirements of the Suggested Licenses still apply. Copilot's failure to provide the attributions for outputs that are modifications of Licensed Materials represents another enormous set of license breaches and DMCA violations.

#### F. Copilot Reproduces the Code of the Named Plaintiffs Without Attribution

97. Because Copilot was trained on all available public GitHub repositories, if Licensed Materials have been posted to a GitHub public repository, Plaintiffs and the Class can be reasonably certain it was ingested by Copilot and is sometimes returned to users as Output.

98. Described below are some specific examples of Copilot's unlawful behavior using Licensed Materials owned by the named Plaintiffs. These examples were emitted by Copilot after prompting Copilot.

99. In the examples below, original code is shaded gray, prompts to Copilot are shaded orange, and outputs from Copilot are shaded light blue.

##### 1. Example: Copilot Outputs the Code of Doe 2 Essentially Verbatim

100. The first example demonstrates Copilot suggesting an essentially verbatim copy of code written by Doe 2.

101. [REDACTED]

[REDACTED] subject to the GNU General Public License v3.0. [REDACTED]

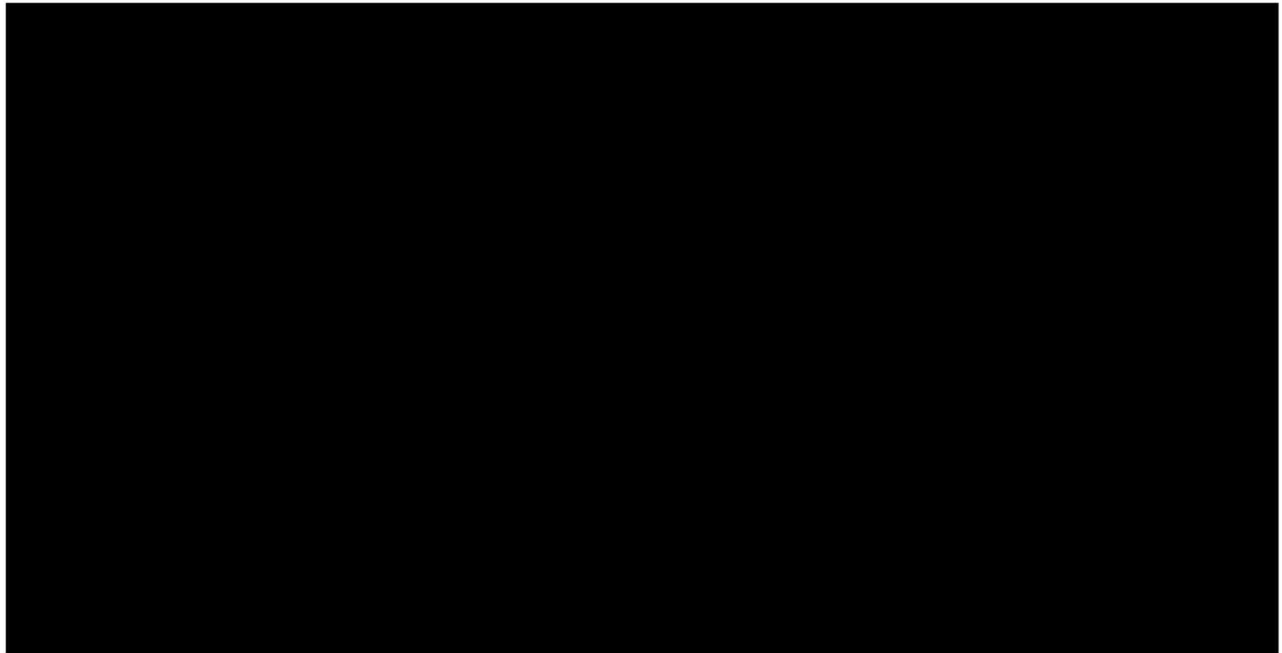
[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

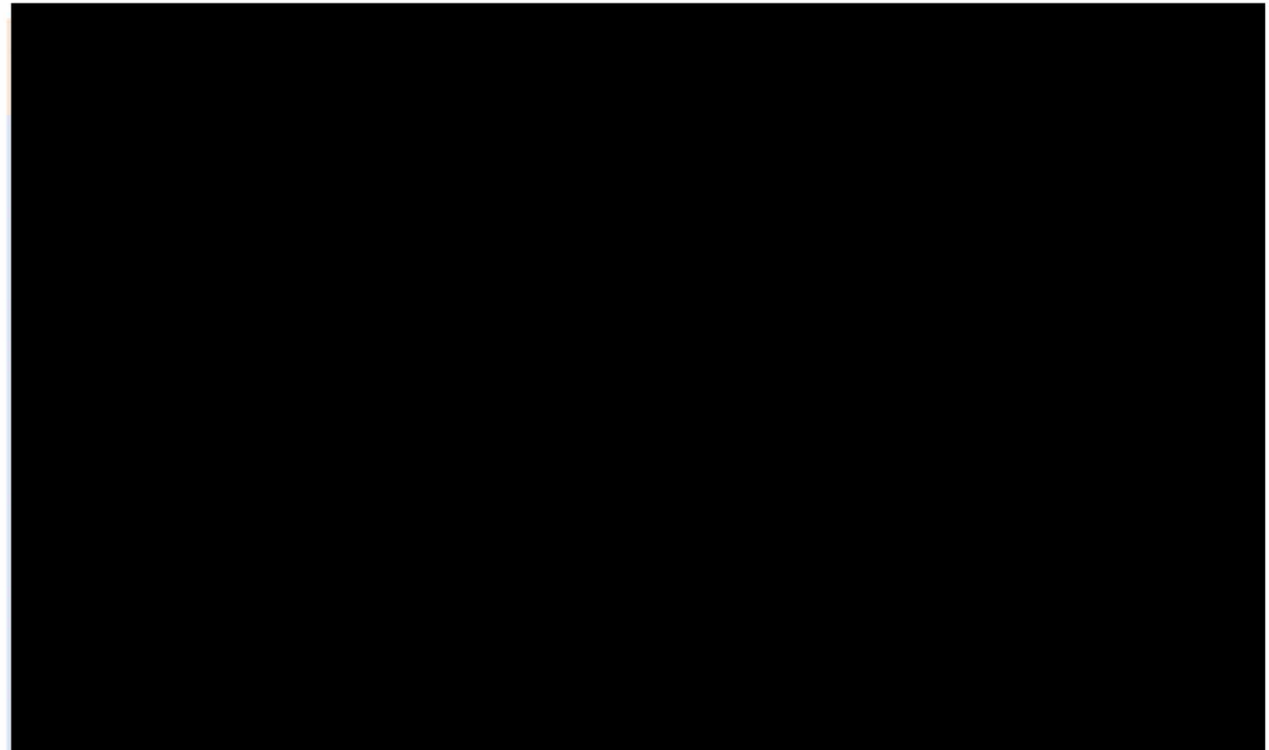




102. When Copilot is prompted the first few lines of Doe 2's code:



Copilot suggests the following:



1           103. This suggestion from Copilot is identical to Doe 2's code, except that [REDACTED]

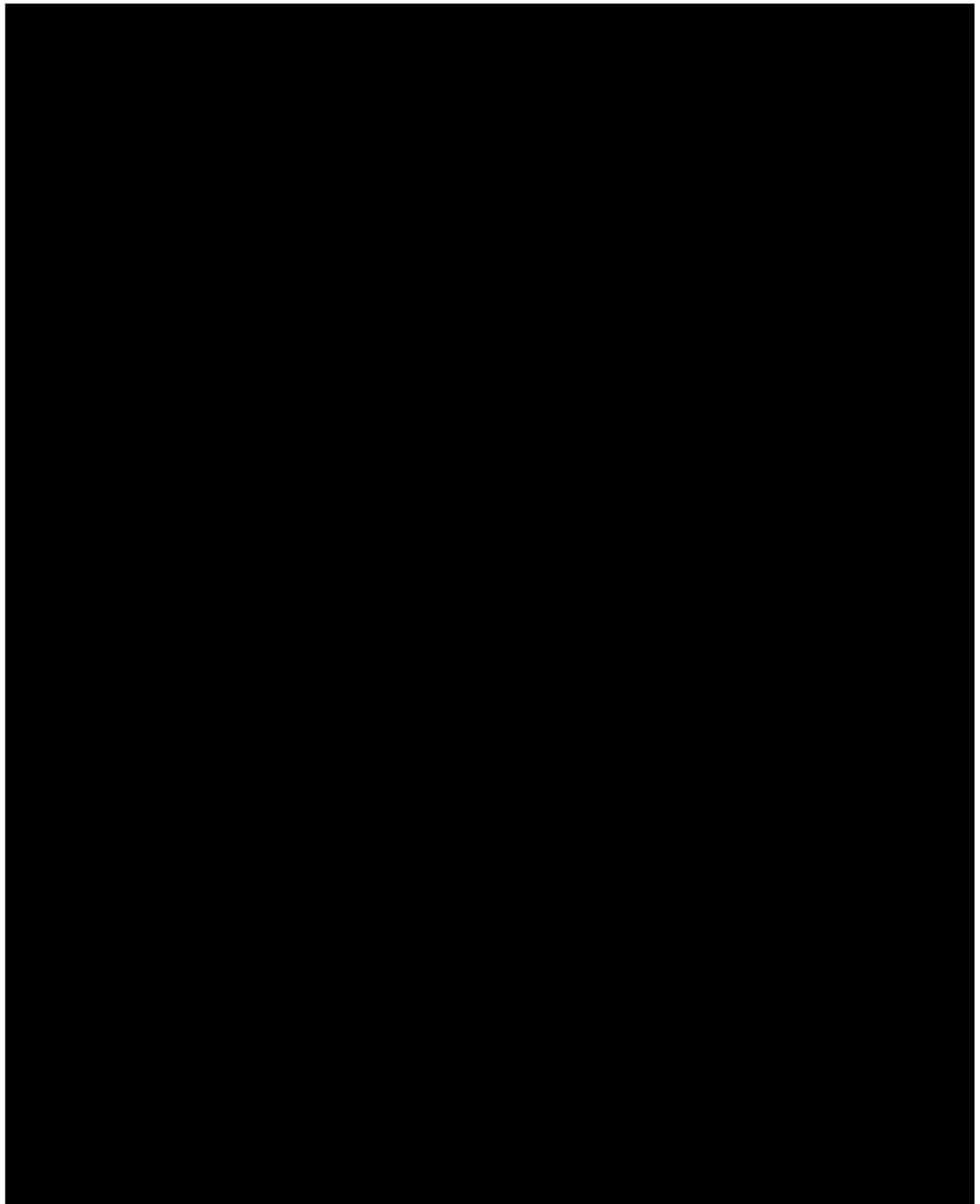
2 [REDACTED]  
3 [REDACTED]  
4 [REDACTED] These differences in the code are cosmetic and the code is  
5 functionally equivalent; otherwise, this is a verbatim copy. Doe 2's particular arrangement and  
6 sequencing seen in his code is distinctive expression found only in one location on GitHub: [REDACTED]  
7 [REDACTED]

8           104. Because the Copilot suggestion is a nearly verbatim reproduction of Doe 2's  
9 unique code, it follows that Copilot copied Doe 2's code. Copilot therefore needed to adhere to  
10 the requirements of Doe 2's license (GNU General Public License v3.0) for that code, including  
11 providing attribution. It does not. Copilot also did not reproduce Doe 2's license.

12                   **2. Example: Copilot Outputs the Code of Doe 1 in Modified Format**

13           105. The second example demonstrates Copilot suggesting a modified copy of code  
14 written by Doe 1. To protect Doe 1's identity, the paragraphs describing the code will be redacted.

15           106. [REDACTED]  
16 subject to the MIT License. [REDACTED]  
17 [REDACTED]  
18 [REDACTED]  
19 [REDACTED]  
20 [REDACTED]  
21 [REDACTED]  
22 [REDACTED]  
23 [REDACTED]  
24 [REDACTED]  
25 [REDACTED]  
26 [REDACTED]  
27 [REDACTED]  
28 [REDACTED]

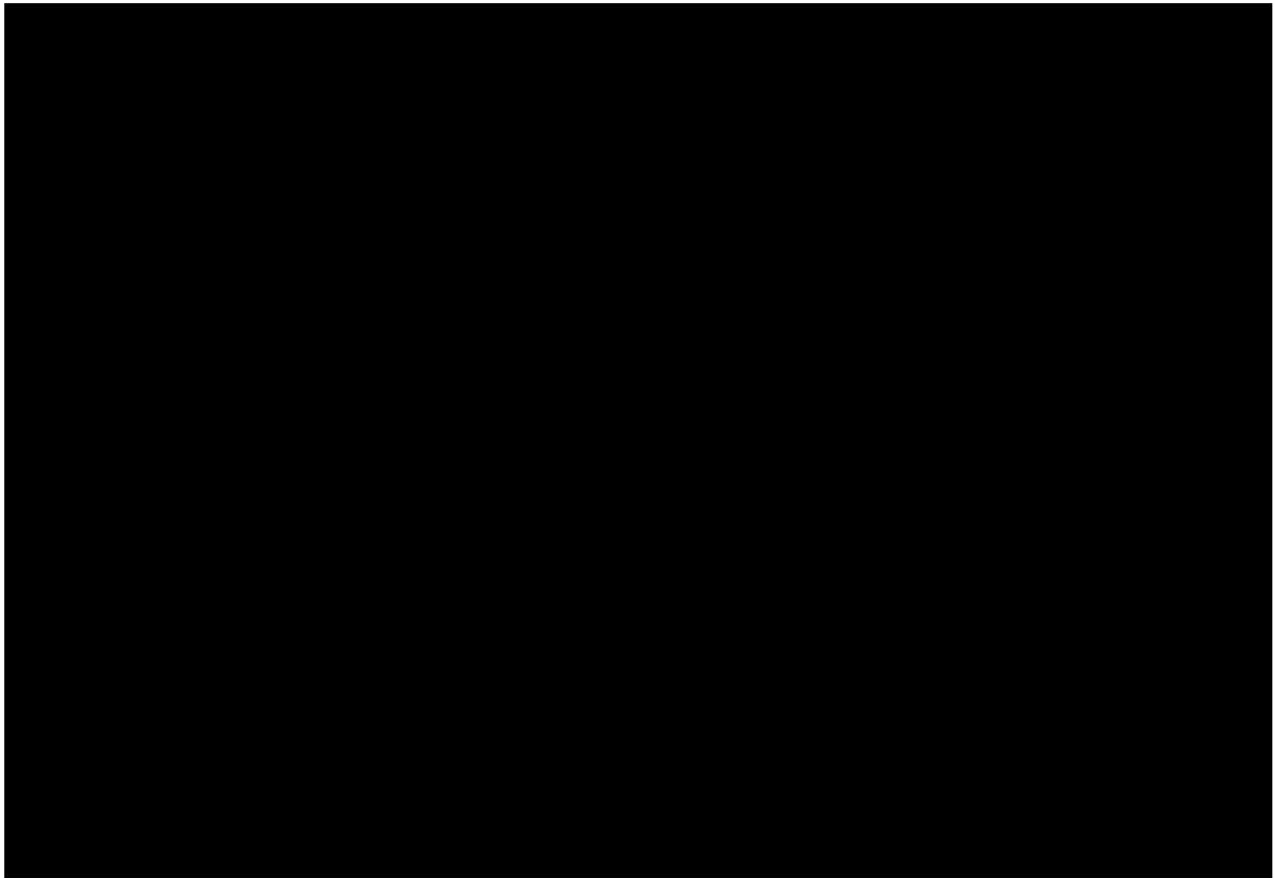


107. When Copilot is prompted with [REDACTED]

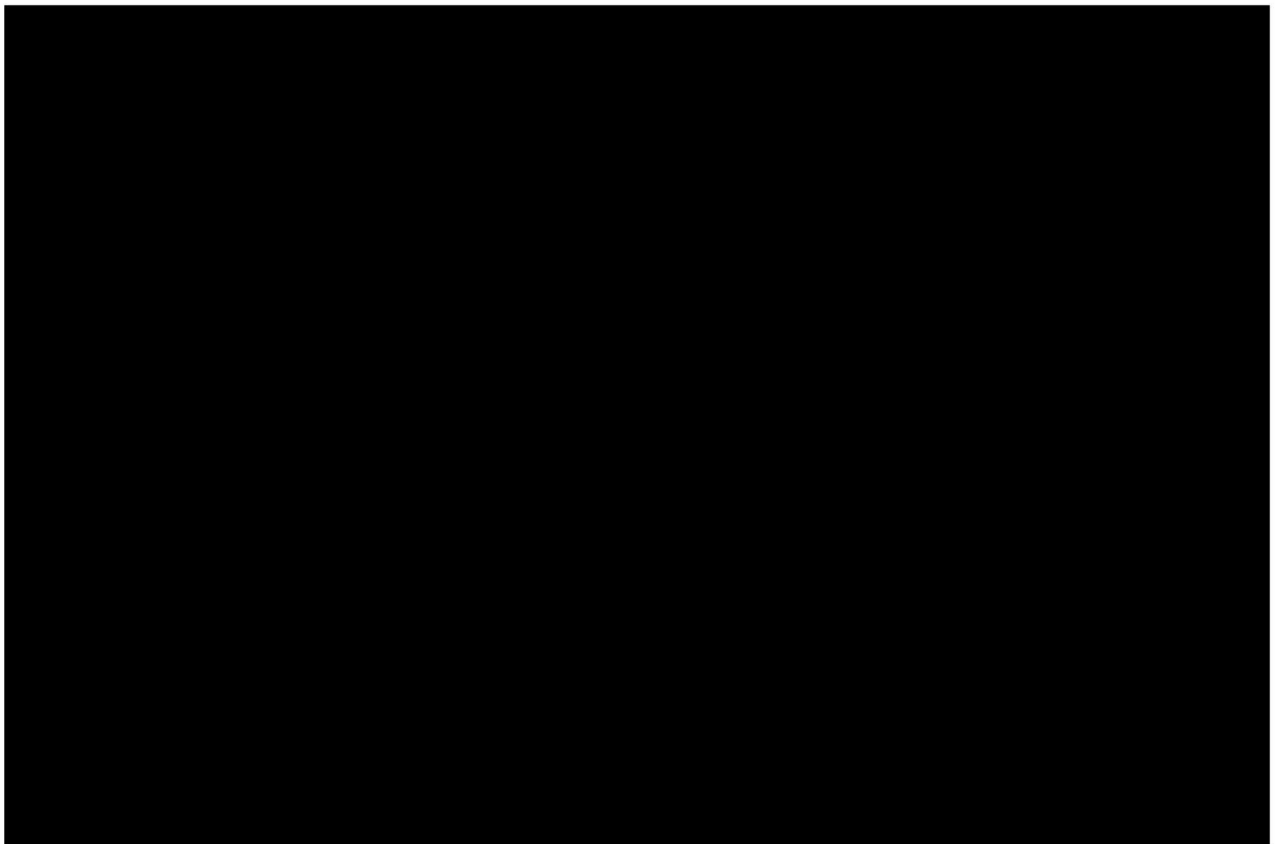
[REDACTED]

[REDACTED]





The first suggestion from Copilot is a modification of Doe 1's code:



[REDACTED]

108. [REDACTED]

[REDACTED] do not appear in any other source file on GitHub. The only way Copilot knows how to make this suggestion is because it ingested Doe 1's source file as training data. Though the Copilot suggestion is not an exact match for Doe 1's code, it is necessarily a modification based on a copy of Doe 1's code.

109. Furthermore, many distinctive expressive features of Doe 1's code have been preserved in Copilot's suggestion. For instance, Doe 1's comments in the code (in green) are reproduced almost verbatim. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

means the same thing as this Copilot-suggested code:

[REDACTED]

1           110. As is apparent from a cursory glance of this example, the variations between  
2 Copilot's emitted output and Doe 1's source code are cosmetic and the code is functionally  
3 equivalent; it follows that Copilot's output is a copy of Doe 1's code.

4           111. That said, Copilot also introduces mistakes into the code. For instance, [REDACTED]  
5 [REDACTED] [REDACTED] [REDACTED] [REDACTED]  
6 [REDACTED]

7           112. Still, because Copilot is reproducing Doe 1's algorithm in modified format, and the  
8 obligations in Doe 1's license (the MIT License) carry with the code even if the underlying code  
9 is modified, the Copilot suggestion needs to follow the requirements of Doe 1's license for that  
10 code, including providing attribution. It does not. Copilot also did not reproduce Doe 1's license.

11           **3. Example: Copilot Outputs the Code of Doe 5 In Modified Format**

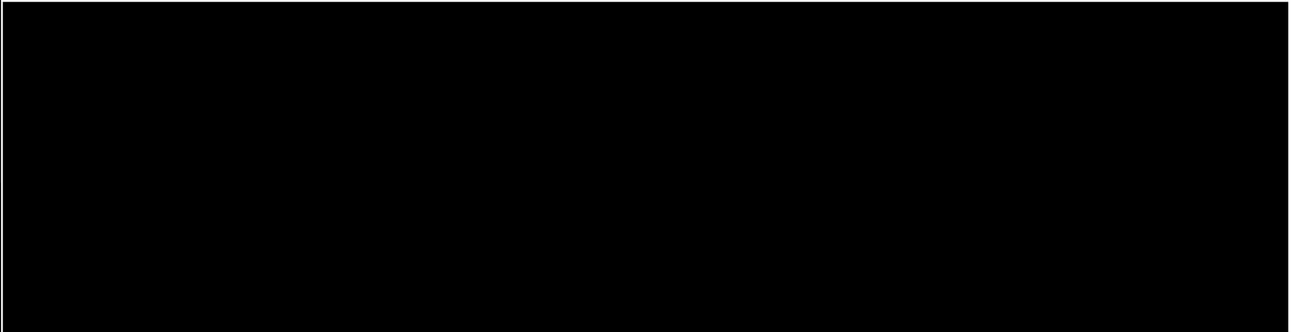
12           113. The third example demonstrates Copilot suggesting multiple modified copies of  
13 code written by Doe 5 in response to a sequence of prompts, which is a common way of using  
14 Copilot. To protect Doe 5's identity, the paragraphs describing the code will be redacted.

15           114. [REDACTED]  
16 [REDACTED] subject to the MIT License. [REDACTED]  
17 [REDACTED] The relevant code from the original source file is shown below:

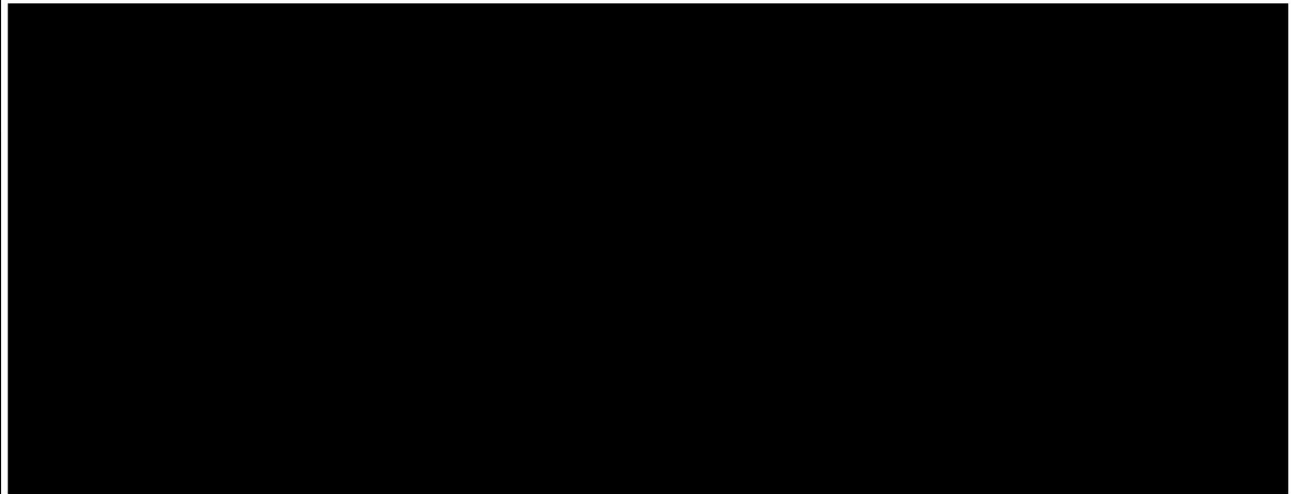
18 [REDACTED]  
19 [REDACTED]  
20 [REDACTED]  
21 [REDACTED]  
22 [REDACTED]  
23 [REDACTED]  
24 [REDACTED]  
25 [REDACTED]  
26 [REDACTED]  
27 [REDACTED]  
28 [REDACTED]



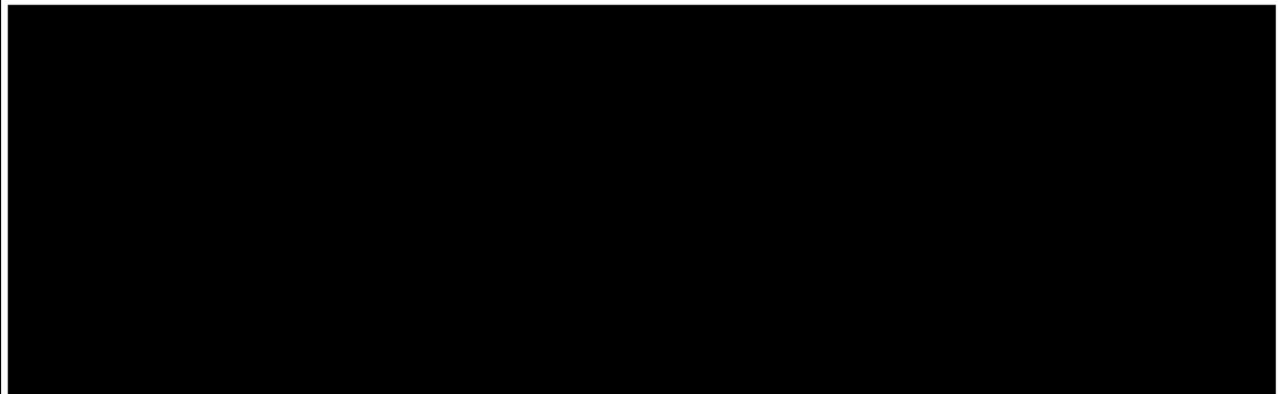
1           115. When Copilot is prompted the first section of Doe 5's code, comprising the first  
2 complete test and the name of the second:



9           116. The first suggestion from Copilot offers to complete the prompt with a verbatim  
10 copy of Doe 5's original code, except that [REDACTED] [REDACTED] [REDACTED]  
11 [REDACTED] (a variation that does not affect how the code works):



20           117. Next, if the name of the third test is appended, the next prompt to Copilot looks  
21 like this:



[REDACTED]

118. The first suggestion from Copilot offers to complete the prompt with a functionally identical copy of Doe 5's code, except [REDACTED] [REDACTED] [REDACTED] [REDACTED] (neither of these variations affect how the code works):

[REDACTED]

119. As is apparent from the high degree of similarity and minor cosmetic deviations between Copilot's emitted output and Doe 5's source code, Copilot ingested, copied and reproduced Doe 5's source code as output.

120. Because Copilot is (repeatedly) reproducing Doe 5's original code in modified format, and the obligations in Doe 5's license (the MIT License) carries with the code even when it is modified, the Copilot suggestions need to follow the requirements of Doe 5's license for that code, including providing attribution. They do not. Copilot also did not reproduce Doe 5's license.

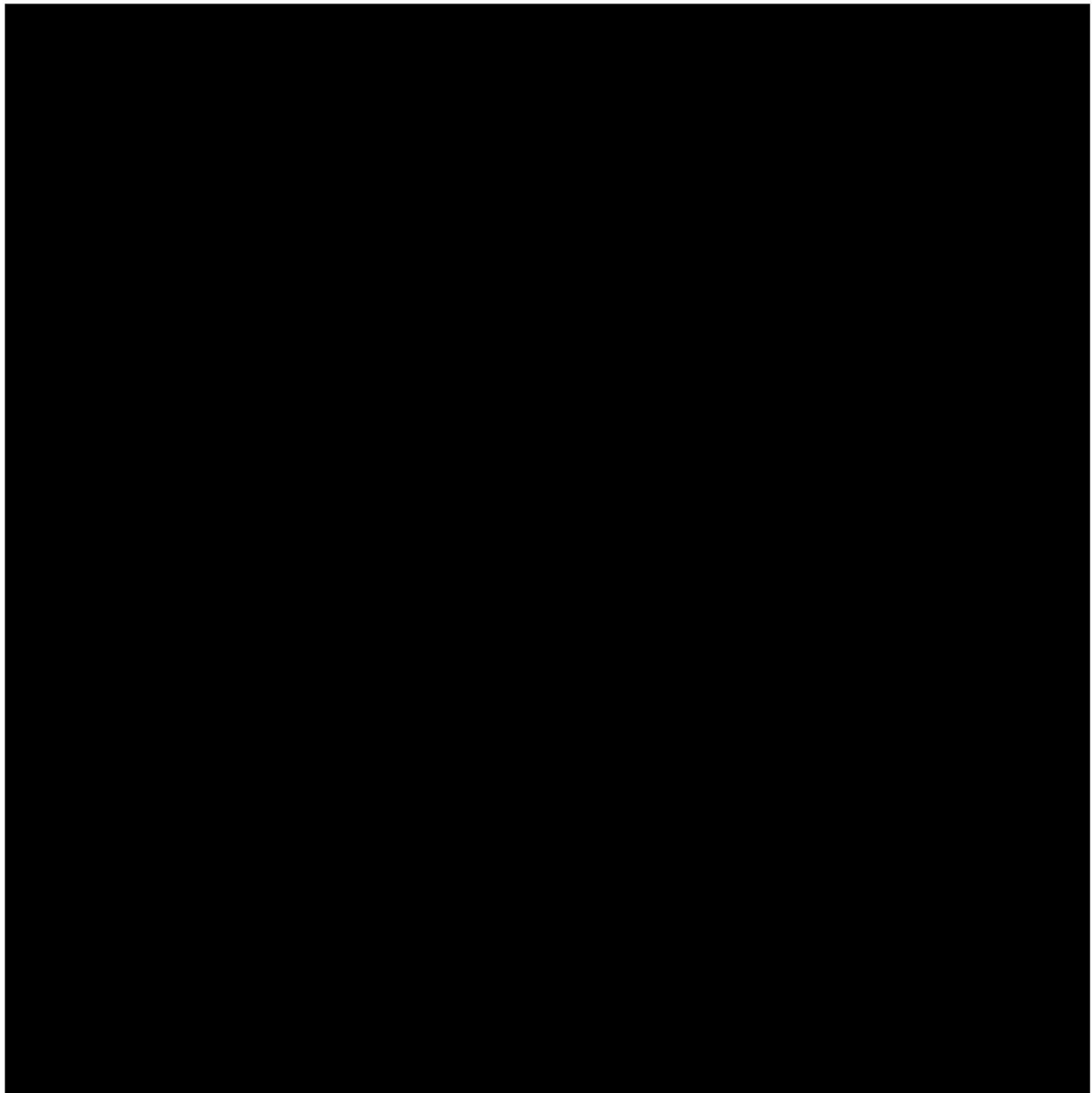
**4. Example: Copilot Outputs Code of Doe 5 Essentially Verbatim**

121. The fourth example also demonstrates Copilot suggesting multiple modified copies of code written by Doe 5 in response to a sequence of prompts, which is a common way of using Copilot. To protect Doe 5's identity, the paragraphs describing the code will be redacted.

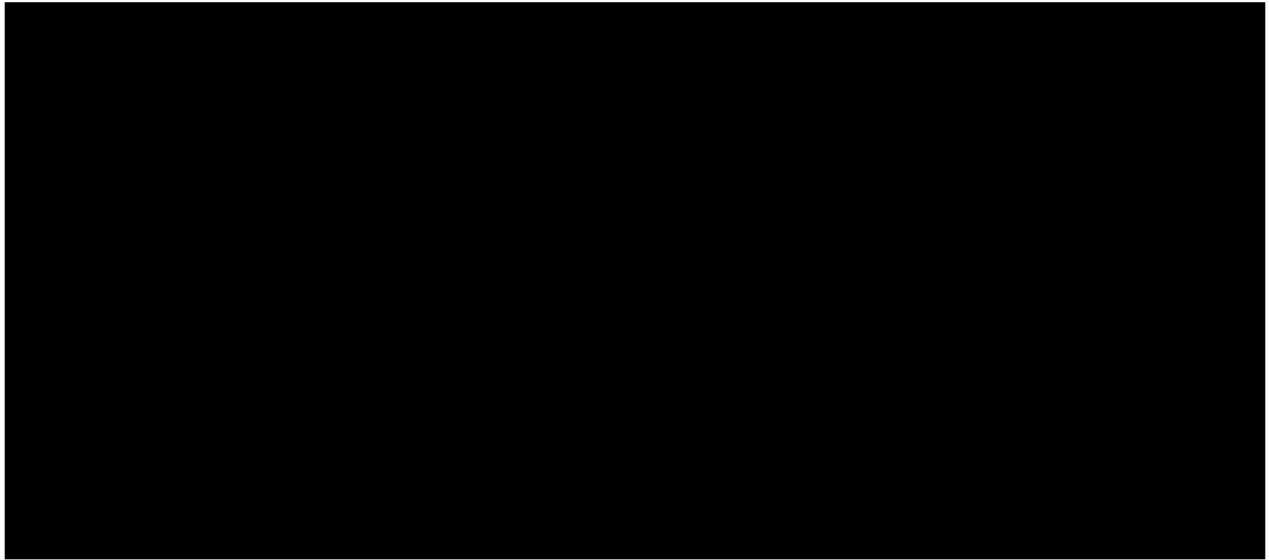
122. [REDACTED]

[REDACTED] subject to the MIT License. [REDACTED]

[REDACTED]. The first three tests from the original source file are shown below:



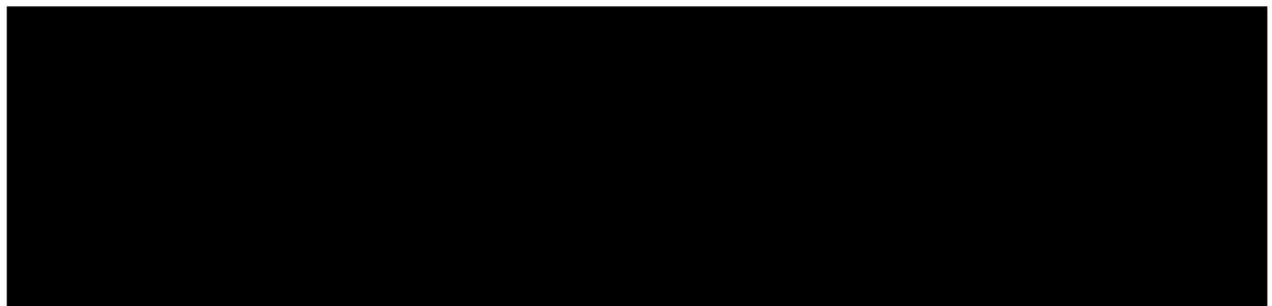


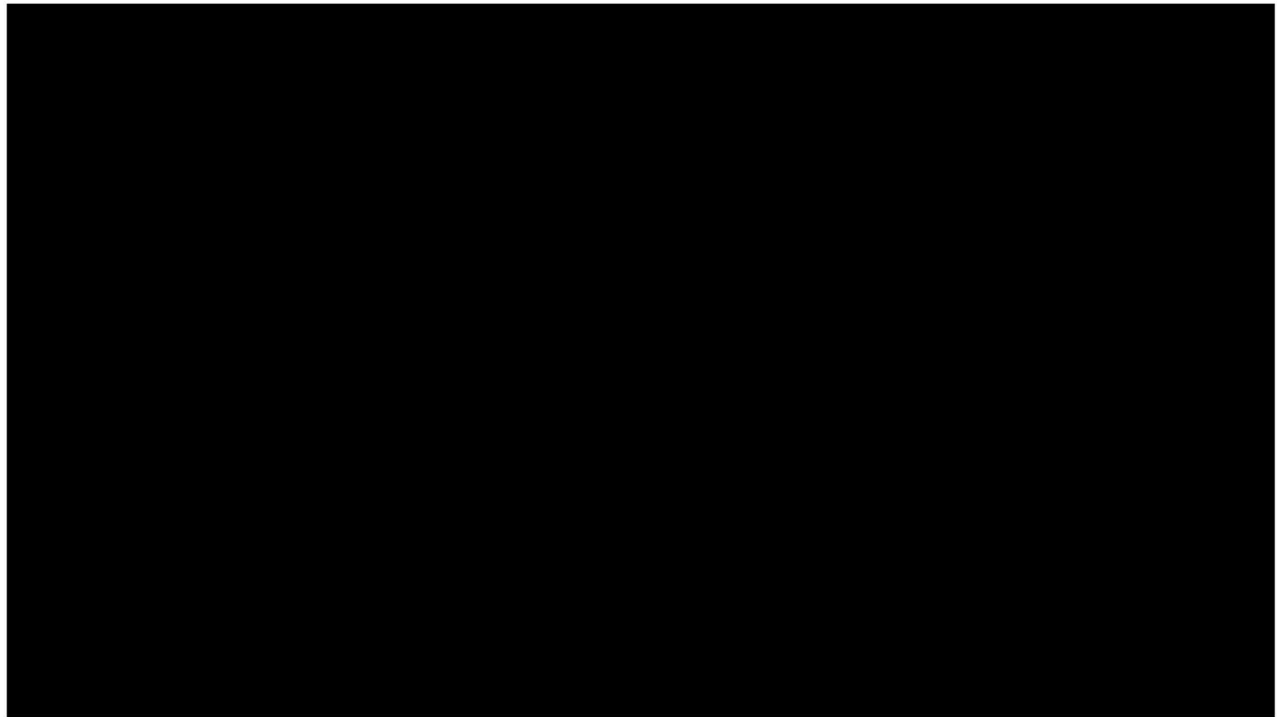


123. When Copilot is prompted with the first section of Doe 5's code, comprising the first complete test and the name of the second:

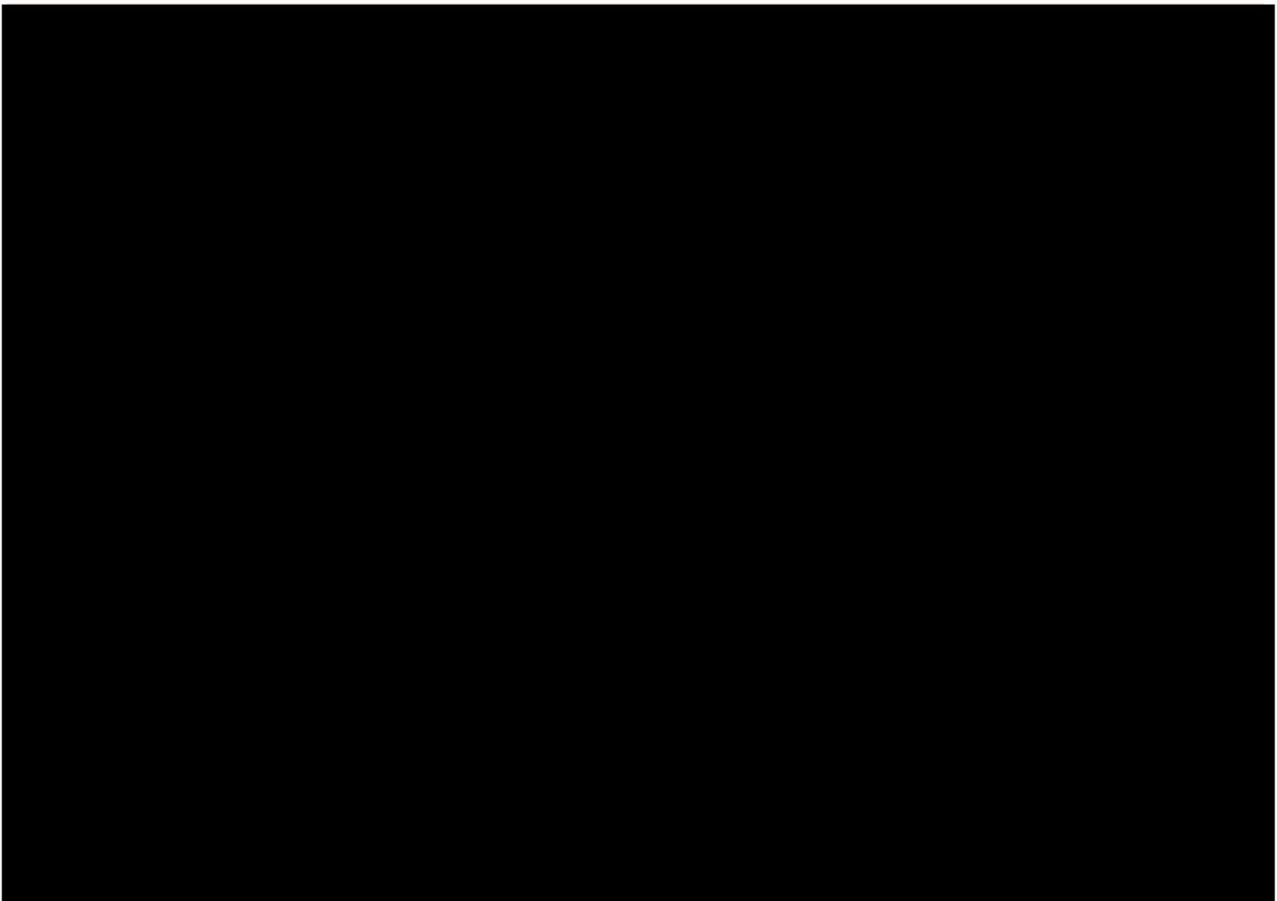


The first suggestion from Copilot offers to complete the second test with a verbatim copy of Doe 5's original code:



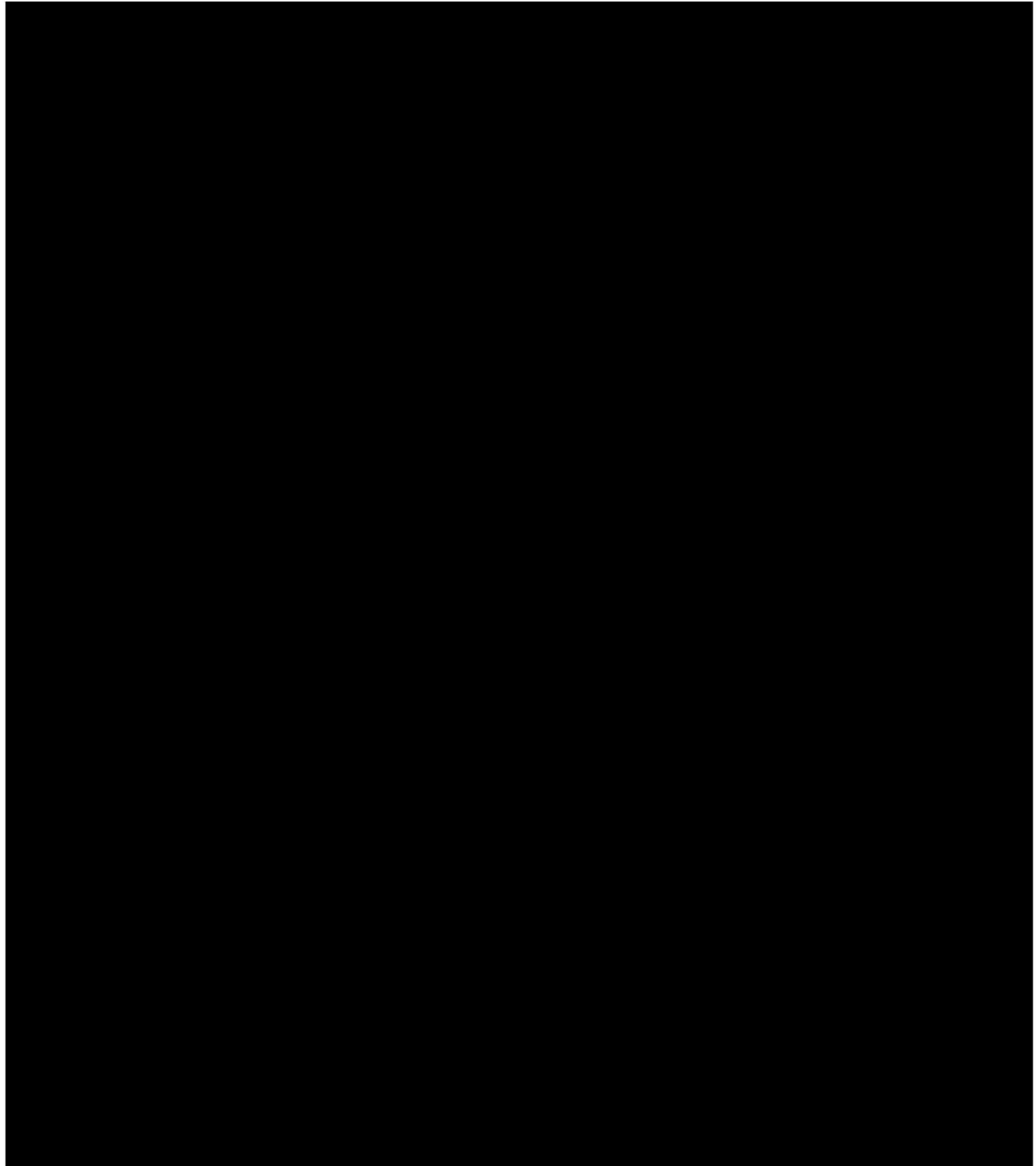


124. When Copilot's suggestion is accepted and the name of Doe 5's third test is appended, the next prompt to Copilot looks like this:





125. Once again, the first suggestion from Copilot offers to complete the third test with a verbatim copy of Doe 5's code (except for small cosmetic variations in line breaks):





126. Because Copilot is (repeatedly) reproducing Doe 5’s code essentially verbatim, the Copilot suggestions need to follow the requirements of Doe 5’s license (the MIT License) for that code, including providing attribution. They do not. Copilot also did not reproduce Doe 5’s license.

127. These are only a few examples of Plaintiffs’ code being reproduced by Copilot. It follows that many if not all prompts entered into Copilot will readily cause it to emit verbatim, near-verbatim or modified copies of Licensed Material that violate the licenses under which the source code is published. Multiplied across the many users of Copilot and the many times Copilot is prompted, each day these violations must be accruing with astonishing frequency. It is therefore likely if not certain that verbatim, near-verbatim or modified copies of each Plaintiffs’ code have already been emitted by Copilot.

128. Additionally, even though Plaintiffs have been able to generate these examples, Plaintiffs remain at a great evidentiary disadvantage relative to Defendants, because Defendants control all the information about the training dataset. In particular, only Defendants know *when* the Licensed Materials of Plaintiffs and the Class were scraped. As is typical in open source, many of the Licensed Materials are regularly updated. As such, it is difficult to determine which iterations of code may have been trained on and would be subject to emission by Copilot.

**G. Codex and Copilot Were Designed to Withhold Attribution, Copyright Notices, and License Terms from Their Users**

129. Codex and Copilot have no way to determine whether license text or other Copyright Management Information (“CMI”)<sup>21</sup> is part of the code it appears immediately before

---

<sup>21</sup> CMI is defined in detail below in Paragraph 187.



1 or after. Unless instructed otherwise, it will assume that CMI that usually appears just before a  
2 given block of code is an important part of that code or otherwise necessary for it to function.

3 130. It is a common practice to provide the applicable license text at the top of every  
4 source file in the codebase. The purpose of this practice is to avoid the code from being divorced  
5 from the license. This may occur via “vendoring,” a method of creating a derivative work by  
6 including source files from a copyrighted project directly into another project without following  
7 the terms of the license or providing attribution or a copyright notice. Copilot circumvents this  
8 protective measure to mask the degree of vendoring it engages in.

9 131. Early iterations of Copilot reproduced license text. For example, in a blog post,  
10 GitHub noted “In one instance, GitHub Copilot suggested starting an empty file with something  
11 it had even seen more than a whopping 700,000 different times during training—that was the  
12 GNU General Public License.”<sup>22</sup> Copilot no longer suggests licenses in this way because it has  
13 been altered not to. As GitHub explains: “GitHub Copilot *has* changed to require a minimum file  
14 content. So some of the suggestions flagged here would not have been shown by the current  
15 version.”

16 132. In July 2021, near Copilot’s launch, it would sometimes produce license text,  
17 attribution, and copyright notices. This CMI was not always accurate. Copilot no longer  
18 reproduces these types of CMI, incorrect or otherwise, on a regular basis. It has been altered not  
19 to.

20 133. In July 2022, in response to public criticism of Copilot’s mishandling of Licensed  
21 Materials, GitHub introduced a user-settable Copilot filter called “Suggestions matching public  
22 code.” If set to “block,” this filter claims to prevent Copilot from suggesting verbatim excerpts of  
23 “about 150 characters” that come from Licensed Materials. But even assuming the filter works as  
24 advertised, because it only checks for verbatim excerpts, it does nothing to impede the Outputs  
25 from Copilot that are modifications of Licensed Materials. Thus, as a device for respecting the  
26 rights of Plaintiffs and the Class, it is essentially worthless.

---

27  
28 <sup>22</sup> <https://github.blog/2021-06-30-github-copilot-research-recitation/>.

134. In GitHub’s hands, the propensity for small cosmetic variations in Copilot’s Output is a feature, not a bug. These small cosmetic variations mean that GitHub can deliver to Copilot customers unlimited modified copies of Licensed Materials without ever triggering Copilot’s verbatim-code filter. AI models like Copilot often have a setting called *temperature* that specifically controls the propensity for variation in their output. On information and belief, GitHub has optimized the temperature setting of Copilot to produce small cosmetic variations of the Licensed Materials as often as possible, so that GitHub can deliver code to Copilot users that *works* the same way as verbatim code, while claiming that Copilot only produces verbatim code 1% of the time. Copilot is an ingenious method of software piracy.

135. In December 2022, GitHub launched Copilot for Business. The initial terms of service included one notable extra provision compared to ordinary Copilot: a “Defense of Third Party Claims” that read:

GitHub will defend you against any claim by an unaffiliated third-party that your use of GitHub Copilot misappropriated a trade secret or directly infringes a patent, copyright, trademark, or other intellectual property right of a third party, up to the greater of \$500,000.00 USD or the total amount paid to GitHub for the use of GitHub Copilot during the 12 months preceding the claim. GitHub’s defense obligations do not apply if (i) the claim is based on Code that differs from a Suggestion provided by GitHub Copilot, (ii) you fail to follow reasonable software development review practices designed to prevent the intentional or inadvertent use of Code in a way that may violate the intellectual property or other rights of a third party, or (iii) you have not enabled all filtering features available in GitHub Copilot.

136. If Copilot had been designed to reproduce the attribution, license terms, and copyright notices of the Licensed Materials, this kind of contractual reassurance wouldn’t be necessary. With this provision (since removed), GitHub acknowledged that Copilot disrupts—possibly with legal consequences—the relationship between authors and users of open-source software.

#### **H. Open-Source Licenses Began to Appear in the Early 1990s**

137. In 1991, software engineer Linus Torvalds began a project to create a UNIX-like operating system that would run on common PC hardware. This project became known as Linux.

1           138. To encourage adoption of his system, and persuade other programmers to  
2 contribute, he released Linux under what was then an unusual software license called the GNU  
3 General Public License, or GPL.

4           139. The GPL is a software license. But whereas most software licenses required  
5 payment, software under the GPL is provided for free. Whereas most software licenses did not  
6 include source code, GPL software always included source code. And whereas most software  
7 licenses prohibited derivative works, the GPL not only allowed it, but encouraged it.

8           140. In certain ways, however, the GPL still operated like a traditional software license.  
9 For example, consistent with copyright law, it depended on an assertion of copyright by the  
10 software author. Even though GPL software was available at no charge, the GPL contained  
11 conditions on its users as licensees.

12           141. One license requirement was that a program derived from GPL software had to  
13 redistribute certain information about that software:

14                   You may copy and distribute verbatim copies of the Program's  
15                   source code as you receive it, in any medium, provided that you  
16                   conspicuously and appropriately publish on each copy an  
17                   appropriate copyright notice and disclaimer of warranty; keep  
18                   intact all the notices that refer to this General Public License and to  
19                   the absence of any warranty; and give any other recipients of the  
20                   Program a copy of this General Public License along with the  
21                   Program.<sup>23</sup>

19 Failure to adhere to these conditions constituted a violation of the license, triggering the  
20 possibility of legal action. Provisions of the GPL are enforceable, and many GPL licensors have  
21 sought to enforce GPL licenses through court proceedings and other litigation.

22           142. The early years of Linux paralleled the early years of the World Wide Web. The  
23 fact that Linux was free and ran on common computer hardware made it a popular choice for web  
24 servers. Because of its contrarian GPL licensing, Linux became hugely popular. A large ecosystem  
25 of other programs and tools grew around it. This contributed to the explosive growth of the web  
26 and other network services across the rest of the 1990s.

---

27  
28 <sup>23</sup> <https://www.gnu.org/licenses/old-licenses/gpl-1.0.en.html>.



143. In turn, the growth of the World Wide Web made it easier for developers in different places to collaborate on software. The GPL, and licenses like it, were a natural fit for this kind of collaborative work.

144. Around 1998, a new name was coined as an umbrella term for these principles of software licensing and development: *open source*.

# **I. Microsoft Has a History of Flouting Open-Source License Requirements**

145. During the 1980s and 1990s, Microsoft was primarily a software company, focusing largely on operating systems and related applications. These included its DOS operating system and later, its Windows operating system. Windows generated billions of dollars in revenue from its sale and licensing as proprietary software for desktop computers and servers. Microsoft derived substantial income from sale of licensed products and devotes substantial resources to protecting and enforcing such licenses.

146. Windows is a graphical operating system. It allows users to view and store files, run software and games, play videos, and provides a way to connect to the internet.

147. Linux represented a competitive threat to Windows. It ran on the same hardware. It performed many of the same functions. It was free. Many programmers at the time considered Linux to be functionally superior to Windows.

148. Microsoft has engaged in a problematic practice known as “vaporware,” where products are announced but are in fact late, never manufactured, or canceled. Typically the company promising vaporware never has any intention of providing it. The term vaporware was coined by Microsoft in 1982 in reference to the development of its Xenix operating system.

149. Microsoft described its anti-Linux strategy as “FUD,” standing for fear, uncertainty, and doubt. Microsoft focused extra attention to Linux’s open-source aspects.

150. In 1998, a source at Microsoft leaked what became known as the “Halloween Documents”, revealing Microsoft’s thinking on how to counter the competitive threat from Linux. Among other things, the documents emphasized the importance of countering the “long



term developer mindshare threat”, and concluded that to defeat open source, “[Microsoft] must target a process rather than a company.”<sup>24</sup>

151. In 2001, Microsoft CEO Steve Ballmer said “The way the [GPL] is written, if you use any open-source software, you must make the rest of your software open source. . . . Linux is a cancer that attaches itself in an intellectual property sense to everything it touches.”<sup>25</sup> Ballmer’s summary of GPL licensing was not accurate. In 2001, Linux was being used by corporations of every size. The growth of open source up to that point, and since, has been made possible by the open-source community’s respect for and compliance with applicable licenses.

152. In 2001, Microsoft was the defendant in a major software-related antitrust case, *United States v. Microsoft Corporation*.<sup>26</sup> In this case, the U.S. Department of Justice accused Microsoft of maintaining a software monopoly by illegally imposing technical restrictions on manufacturers of personal computers, including “tying” violations related to the Internet Explorer web browser. Judge Thomas Penfield Jackson, who presided over the antitrust trial, opined that Microsoft is “a company with an institutional disdain for both the truth and for rules of law that lesser entities must respect. It is also a company whose ‘senior management’ is not averse to offering specious testimony to support spurious defenses to claims of its wrongdoing.”<sup>27</sup>

153. In 2007, Microsoft admitted that it tried to influence the vote of an ISO open-standards committee by offering money to certain business partners in Sweden to vote for Microsoft’s preferred outcome.<sup>28</sup>

154. After observing the rapid growth of Amazon’s original cloud computing products, Microsoft has expanded its business into cloud computing, which it has branded Microsoft Azure or simply Azure. Microsoft announced Azure to developers in 2008. It was formally released in

---

<sup>24</sup> <http://www.catb.org/esr/halloween/halloween1.html>.

<sup>25</sup> <https://lwn.net/2001/0607/a/esr-big-lie.php3>.

<sup>26</sup> No. Civ.A. 00-1457 TPJ.

<sup>27</sup> *Jackson v. Microsoft Corp.*, 135 F. Supp. 2d 38 (D.D.C. 2001).

<sup>28</sup> <https://learn.microsoft.com/en-us/archive/blogs/jasonmatusow/open-xml-the-vote-in-sweden/>.

2010. Azure uses large-scale virtualization at Microsoft data centers and offers many hundreds of services, including infrastructure as a service (“IaaS”), platform as a service (“PaaS”), compute services, Azure Active Directory, mobile services, storage services, communication services, data management, messaging, developer services, Azure AI, blockchain, and others.

**J. GitHub Was Designed to Cater to Open-Source Projects**

155. By 2002, Linux had become immensely popular. But the project itself had become unwieldy and had outgrown its reliance on informal systems of managing software source code (also known as *source-control systems*). The Linux community needed something better.

156. Linus Torvalds set about writing a new source-control system. He named his new system Git. He released it under the GPL. It quickly became the source-control system of choice for open-source programmers.

157. A single software project stored in Git is called a *source repository*, commonly shortened to *repository* or just *repo*. A Git source repository would typically be stored on a networked server accessible to a group of programmers.

158. This became less convenient, however, when programmers were distributed among multiple locations, rather than being in a single location. A Git repository could be stored on an internet-accessible server. But setting up that server hardware and being responsible for it was inconvenient and expensive.

159. In 2008, a group of open-source developers in San Francisco, California founded GitHub. GitHub managed internet servers that hosted Git source repositories. With an account at GitHub, an open-source developer could easily set up a Git project accessible to collaborators anywhere in the world. From early on, GitHub’s core market was open-source developers, whom it attracted by making many of its hosting services free.

160. Most open-source programmers used GitHub to create “public” repositories, meaning that anyone could view them & access them. GitHub also allowed programmers and organizations to create “private” repositories, which were not accessible from the public GitHub website, and required password access.

1           161. Open-source licensing was integral to GitHub. GitHub encouraged open-source  
2 developers to understand and use open-source licenses for their work. Many—though not all—  
3 public repositories on GitHub carry an open-source license. By convention, this license is stored  
4 at the top level of each repository in a file called LICENSE. GitHub’s interface also includes a  
5 button on the front pages of most repositories users can click to see details of the applicable  
6 license. A human user could easily find the license in either of these locations—as could an AI  
7 anywhere near as powerful as Codex or Copilot.

8           162. Though the GPL is one of the early open-source licenses and remains common, it  
9 is not the only open-source license. Examples of other common open-source licenses include the  
10 MIT License, the Apache License, and the Berkeley Software Distribution License (all of which  
11 are included in the Suggested Licenses).

12           163. Though these licenses differ in their wording and their details, most of them share  
13 a requirement that a copy of the license be included with any copy, derivative, or redistribution of  
14 the software, and that the author’s name and copyright notice remains intact. This is not a  
15 controversial requirement of open-source licenses—indeed, it has been an integral part of the  
16 GPL for over 30 years.

17           164. There are also many public repositories on GitHub that have no license. Though  
18 GitHub has encouraged awareness of licenses among its users, it has never imposed a default  
19 license on public repositories. A public repository without a license is subject to ordinary rules of  
20 U.S. copyright.

21           165. Open-source developers flocked to GitHub. By 2018, GitHub had become the  
22 largest and most successful Git hosting service, hosting millions of users and projects.

23           166. In October 2018, Microsoft acquired GitHub for \$7.5 billion. It was important to  
24 Microsoft that programmers use GitHub. Microsoft had developed a well-deserved poor  
25 reputation because of its documented vaporware, FUD, and other business practices, including  
26 those targeted at open-source programs and programming, and open-source licensing specifically.  
27 Microsoft made false and misleading statements and omissions to assuage such concerns,  
28



1 including its primary mantra intended to win over the open-source community: “Microsoft Loves  
2 Open Source.”

3 **K. OpenAI Is Intertwined with Microsoft and GitHub**

4 167. OpenAI, Inc. is a nonprofit corporation founded in December 2015 by a group that  
5 included Greg Brockman, Ilya Sutskever, and other AI researchers; Elon Musk, CEO of Tesla;  
6 and Sam Altman, president of Y Combinator, a tech-startup incubator with hundreds of  
7 companies in its portfolio. Musk and Altman served as co-chairs of OpenAI, Inc. One of OpenAI,  
8 Inc.’s current board members is Reid Hoffman, founder of LinkedIn, which is now a Microsoft  
9 subsidiary. Mr. Hoffman is also a member of the Microsoft Board of Directors.

10 168. Less than a year later, in November 2016, OpenAI first partnered with Microsoft.  
11 It described the partnership as follows: “We’re working with Microsoft to start running most of  
12 our large-scale experiments on Azure. This will make Azure the primary cloud platform that  
13 OpenAI is using for deep learning and AI, and will let us conduct more research and share the  
14 results with the world.”

15 169. Initially, OpenAI, Inc. held itself out as a “non-profit artificial intelligence research  
16 company” that sought to shape AI “in the way that is most likely to benefit humanity as a whole.”

17 170. OpenAI, Inc. reportedly secured \$1 billion in initial funding, from sources that  
18 were largely not disclosed, but included at least most of its founders.

19 171. OpenAI, Inc. obtained its initial source of training data from its founders’  
20 companies. According to reporting at the time, Musk and Altman planned to “pool[] online data  
21 from their respective companies” to serve as training data for OpenAI, Inc. projects. Musk  
22 planned to contribute data from Tesla; Altman planned to have Y Combinator companies “share  
23 their data with OpenAI.”<sup>29</sup>

24 172. In February 2019, Altman created OpenAI, LP, a for-profit subsidiary of the  
25 nonprofit entity OpenAI, Inc. The new OpenAI, LP entity would serve as a vessel for accepting  
26 traditional outside investment in exchange for equity and distributing profits.

---

27 <sup>29</sup> [https://www.wired.com/2015/12/elon-musks-billion-dollar-ai-plan-is-about-far-more-than-](https://www.wired.com/2015/12/elon-musks-billion-dollar-ai-plan-is-about-far-more-than-saving-the-world/)  
28 [saving-the-world/](https://www.wired.com/2015/12/elon-musks-billion-dollar-ai-plan-is-about-far-more-than-saving-the-world/).



173. In July 2019, OpenAI, L.P. accepted a \$1 billion investment from Microsoft. In addition to cash, Microsoft would become the exclusive licensor of certain OpenAI, LP products (including GPT-3, described below in Paragraph 176). Also, as part of this alliance, OpenAI, LP would use Microsoft’s cloud-computing platform, Azure, exclusively to develop and host its products. Some portion of Microsoft’s investment was paid in credits for use of Azure rather than cash. Finally, Microsoft and OpenAI agreed to “jointly build new Azure AI supercomputing technologies.”

174. Azure is a major growth area for Microsoft. In its most recent earnings report on October 25, 2022, “Azure and other cloud services” grew by 35% from the previous quarter, more than any other product.<sup>30</sup> Azure has grown rapidly since Microsoft began its partnership with OpenAI in 2016. Its revenue grew by 50% or more every quarter from 2016 through the first three quarters of 2020.

175. In May 2020, Microsoft and OpenAI announced they had jointly built a supercomputer in Azure that would be used exclusively by OpenAI to train its AI models. Microsoft’s influence over and frequent collaboration with OpenAI has led some to describe Microsoft as “the unofficial owner of OpenAI.”<sup>31</sup>

176. One of OpenAI’s projects is GPT-3, a so-called “large language model” designed to emit naturalistic text. When researchers noticed that GPT-3 could also generate software code, they started studying whether they could make a new AI model specifically trained for this purpose. This project became known as Codex.

177. Sometime after July 2019, OpenAI and Microsoft began collaborating on a code-completion product for GitHub that would use Codex as its underlying model. This product became known as Copilot.

178. On September 28, 2022, OpenAI released an image-generation AI called DALL-E-2. Much like Copilot, DALL-E-2 removes any attribution and/or copyright notice from the

---

<sup>30</sup> <https://www.microsoft.com/en-us/Investor/earnings/FY-2023-Q1/press-release-webcast/>.

<sup>31</sup> <https://venturebeat.com/ai/what-to-expect-from-openais-codex-api/>.

images it uses to create derivative works. Like with Codex, here, OpenAI ignores the rights of the owners of copyrights to images it has ingested.

179. In another joint project, Microsoft and OpenAI recently launched a preview of a product called “Azure OpenAI Service.”<sup>32</sup> This service will “Leverage large-scale, generative AI models with deep understandings of language and code to enable new reasoning and comprehension capabilities for building cutting-edge applications. Apply these coding and language models to a variety of use cases, such as writing assistance, code generation, and reasoning over data. Detect and mitigate harmful use with built-in responsible AI and access enterprise-grade Azure security.”

#### **L. Conclusion of Factual Allegations**

180. Future AI products may represent a bold and innovative step forward. GitHub Copilot and OpenAI Codex, however, do not. Defendants should not have released these products until they could ensure that they did not constantly violate Plaintiffs’ and the Class’s intellectual-property rights, licenses, and other rights.

181. Defendants have made no attempt to comply with the open-source licenses that are attached to much of their training data. Instead, they have pretended those licenses do not exist, and trained Codex and Copilot to do the same. By simultaneously violating the open-source licenses of tens-of-thousands—possibly millions—of software developers, Defendants have accomplished software piracy on an unprecedented scale. As Microsoft’s Co-Founder Bill Gates once said regarding software piracy: “the thing you do is theft.”<sup>33</sup>

182. There is no inherent limitation or constraint of AI systems that made any of this necessary. Defendants chose to build AI systems designed to enhance their own profit at the expense of a global open-source community that they had once sought to foster and protect. GitHub and OpenAI are profiting at the expense of Plaintiffs’ and the Class’s rights.

---

<sup>32</sup> <https://azure.microsoft.com/en-us/products/cognitive-services/openai-service/>.

<sup>33</sup> [https://www.digibarn.com/collections/newsletters/homebrew/V2\\_01/gatesletter.html](https://www.digibarn.com/collections/newsletters/homebrew/V2_01/gatesletter.html)

**VIII. CLAIMS FOR RELIEF**

**COUNT 1**  
**VIOLATION OF THE DIGITAL MILLENNIUM COPYRIGHT ACT**  
**17 U.S.C. §§ 1201–1205**  
**(Direct, Vicarious, and Contributory)**  
**(Against All Defendants)**

183. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

184. As described herein, Defendants have intentionally removed or altered CMI from Plaintiffs' code in violation of Section 1202(b)(1) of the DMCA.

185. As described herein, Defendants have distributed copies of Plaintiffs' code knowing that CMI has been removed or altered while knowing or having reasonable grounds to know that it will induce, enable, facilitate, or conceal infringement in violation of Section 1202(b)(3) of the DMCA.

186. Plaintiffs and members of the Class own the copyrights to Licensed Materials used to train Codex and Copilot. Copilot was trained on millions—possibly billions—of lines of code publicly available on GitHub. Copilot runs on Microsoft's Azure cloud platform exclusively and Microsoft had input in the creation of Copilot. Microsoft is aware that Copilot ignores License Terms and that it was trained almost exclusively on Licensed Materials.

187. Plaintiffs and members of the Class included the following Copyright Management Information (as defined in Section 1202(c) of the DMCA) ("CMI") in the Licensed Materials:

- a. copyright notices;
- b. the title and other information identifying the Licensed Materials;
- c. the name of, and other identifying information about, the authors of the Licensed Materials;
- d. the name of, and other identifying information about, the copyright owners of the Licensed Materials;
- e. terms and conditions for use of the Licensed Materials, specifically the Suggested Licenses; and



f. identifying numbers or symbols referring to CMI or links to CMI.

188. Defendants did not contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA.

189. Defendants knew that they did not contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA.

190. As part of the scheme, Defendants did not attempt to contact Plaintiffs to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA. In fact, Defendants' removal of CMI made it difficult or impossible to contact Plaintiffs and the Class to obtain authority to remove or alter CMI from the Licensed Materials within the meaning of the DMCA. Rather, Defendants removed or altered CMI from open-source code that is owned by Plaintiffs and the Class after the code was uploaded to a GitHub repository by incorporating it into Copilot with its CMI removed.

191. Without the authority of Plaintiffs and the Class, Defendants intentionally removed or altered CMI from the Licensed Materials after they were uploaded to one or more GitHub repositories.

192. Defendants had access to but were not licensed by Plaintiffs nor the Class to train any machine learning, AI, or other pseudo-intelligent computer program, algorithm, or other functional prediction engine using the Licensed Materials.

193. Defendants had access to but were not licensed by Plaintiffs nor the Class to incorporate the Licensed Materials into Copilot.

194. Defendants had access to but were not licensed by Plaintiffs nor the Class to create Derivative Works<sup>34</sup> based upon the Licensed Materials.

195. Defendants had access to but were not licensed by Plaintiffs nor the Class to distribute the Licensed Materials as they do through Copilot.

---

<sup>34</sup> "Derivative Works" as used herein refers to Copilot's Output to the extent they are derived from Licensed Materials. The definition also includes the Copilot product itself, which is a Derivative Work based upon a large corpus of Licensed Materials.



1           196.   Without the authority of Plaintiffs and the Class, Defendants distributed CMI  
2 knowing that the CMI had been removed or altered without authority of the copyright owner or  
3 the law with respect to the Licensed Materials.

4           197.   Defendants distributed copies of the Licensed Materials knowing and intending  
5 that CMI had been removed or altered without authority of the copyright owner or the law, with  
6 respect to the Licensed Materials.

7           198.   Defendants removed or altered CMI from the Licensed Materials knowing and  
8 intending that it would induce, enable, facilitate, or conceal infringement of copyright.

9           199.   Without the CMI associated with the Licensed Materials, Copilot users are  
10 induced or enabled to copy the Licensed Materials. Because CMI has been removed, Copilot  
11 users do not know whether Output is owned by someone else and subject to restrictions on use.  
12 Without the CMI, copyright infringement is facilitated or concealed, because Plaintiffs and the  
13 Class are prevented from knowing or learning that the Output is based upon one or more of the  
14 Licensed Materials. Use of the Licensed Materials is not infringement when the terms of the  
15 applicable Suggested License are followed. Had the CMI not been removed, Copilot users would  
16 be aware of the Licenses and their obligations under them. The terms of the applicable Suggested  
17 License would have allowed those users to use the Licensed Materials without infringement. By  
18 withholding and concealing license information and other CMI, Defendants prevented Copilot  
19 users from making non-infringing use of the Licensed Materials. This contradicts the express  
20 wishes of Plaintiffs and the Class, which are set forth explicitly in the Suggested Licenses under  
21 which the Licensed Materials are offered.

22           200.   Defendants removed or altered CMI from Licensed Materials owned by Plaintiffs  
23 and the Class while possessing reasonable grounds to know that it would induce, enable, facilitate,  
24 and/or conceal infringement of copyright in violation of Sections 1202(b)(1) and 1202(b)(3) of  
25 the DMCA.

26           201.   By omitting, altering and/or concealing CMI from Copilot's Output, Defendants  
27 have reasonable grounds to know that innocent infringers are induced or enabled to copy the  
28 Licensed Materials, because CMI has been removed. Without the CMI, Defendants have

reasonable grounds to know copyright infringement is facilitated or concealed, because Plaintiffs and the Class have the difficult or impossible task of proving the Licensed Materials belong to them.

202. The profits attributable to Defendants' violation of the DMCA include the revenue from: Copilot subscription fees, sales of or subscriptions to Defendants' Copilot-related products and/or services that are used to run Copilot, hosting Copilot on Azure, and any other of Defendants' products that contain copies of the Licensed Materials without all the original CMI. The Licensed Materials add nearly all value to the Copilot product because the purpose of Copilot is to provide code and the source of that code is the Licensed Materials. Without the Licensed Materials, Copilot would not be functional.

203. On information and belief, Defendants could have trained Copilot to include attribution, copyright notices, and license terms when it provides Output covered by a License.

204. Defendants did not request or obtain permission from Plaintiffs and the Class to use the Licensed Materials for Defendants' Copilot product.

205. Defendants use of the Licensed Materials does not follow the requirements of the Suggested Licenses associated with the Licensed Materials. In particular, Copilot fails to provide attribution for the creator nor the owner of the Work. Copilot fails to include the required copyright notice included in the License. Copilot fails to include the applicable Suggested License's text.

206. Defendants are sophisticated with respect to intellectual property matters related to open-source code. Microsoft in particular has extensive experience granting licenses, obtaining licenses, and enforcing license terms. Its most recent Annual Report states:

**We protect our intellectual property investments in a variety of ways. We work actively in the U.S. and internationally to ensure the enforcement of copyright, trademark, trade secret, and other protections that apply to our software and hardware products, services, business plans, and branding.** We are a leader among technology companies in pursuing patents and currently have a portfolio of over 69,000 U.S. and international patents issued and over 19,000 pending worldwide. While we employ much of our internally-developed intellectual property exclusively in our products and services, we also engage in

1 outbound licensing of specific patented technologies that are  
2 incorporated into licensees' products. From time to time, we enter  
3 into broader cross-license agreements with other technology  
4 companies covering entire groups of patents. We may also purchase  
5 or license technology that we incorporate into our products and  
6 services. At times, we make select intellectual property broadly  
7 available at no or low cost to achieve a strategic objective, such as  
8 promoting industry standards, advancing interoperability,  
9 supporting societal and/or environmental efforts, or attracting and  
10 enabling our external development community. **Our increasing  
11 engagement with open source software will also cause us to  
12 license our intellectual property rights broadly in certain  
13 situations.**

14 Microsoft Corporation Annual Report, Form 10-K at 27 (July 28, 2022) (emphasis added).<sup>35</sup>

15 207. GitHub, which offers the Copilot product jointly with OpenAI, also has extensive  
16 experience with the DMCA. GitHub knows or reasonably should know that the Licensed  
17 Materials it hosts are subject to copyright. It provides the language of the Suggested Licenses to  
18 users, all of which include copyright notices. Its 2022 Transparency Report—January to June<sup>36</sup>  
19 states: “Copyright-related takedowns (which we often refer to as DMCA takedowns) are  
20 particularly relevant to GitHub because so much of our users’ content is software code and can be  
21 eligible for copyright protection.”<sup>37</sup> In the first six months of 2022, GitHub processed 1220  
22 DMCA takedown requests. Its DMCA Takedown Policy<sup>38</sup> notes “GitHub probably never would  
23 have existed without the DMCA.”

24 208. GitHub also knows or reasonably should know the portions of the DMCA giving  
25 rise to Plaintiffs’ claim. In its 2021 Transparency Report, “Before removing content based on  
26 alleged circumvention of copyright controls (under Section 1201 of the US DMCA or similar laws  
27 in other countries), we carefully review both the legal and technical claims, and we sponsor a  
28 Developer Defense Fund to provide developers with meaningful access to legal resources.”<sup>39</sup>

---

24 <sup>35</sup> <https://microsoft.gcs-web.com/static-files/07cf3c30-cfc3-4567-b20f-f4b0f0bd5087/>.

25 <sup>36</sup> <https://github.blog/2022-08-16-2022-transparency-report-january-to-june/>.

26 <sup>37</sup> <https://github.blog/2022-08-16-2022-transparency-report-january-to-june/>.

27 <sup>38</sup> <https://docs.github.com/en/site-policy/content-removal-policies/dmca-takedown-policy#what-is-the-dmca/>.

28 <sup>39</sup> <https://github.blog/2022-01-27-2021-transparency-report/>.



209. GitHub is aware that Copilot’s removal of CMI is illegal. For example, it states that “publishing or sharing tools that enable circumvention are not [permitted]”<sup>40</sup> and “Distributing tools that enable circumvention is prohibited, even if their use by developers falls under the exemption [for security research].”<sup>41</sup> GitHub has also frequently published articles discussing the DMCA, its application, and the Copyright Office’s guidance on its scope and exceptions.<sup>42</sup>

210. Unless Defendants are enjoined from violating the DMCA, Plaintiffs and the Class will suffer great and irreparable harm by depriving them of the right to identify and control the reproduction and/or distribution of their copyrighted works, to have the terms of their open-source licenses followed, and to pursue copyright-infringement remedies. Defendants will not be damaged if they are required to comply with the DMCA. Plaintiffs and the Class are therefore entitled to an injunction barring Defendants from violating the DMCA and impounding any device or product that is in the custody or control of Defendants and that the court has reasonable cause to believe was involved in a violation of the DMCA.

211. Plaintiffs and the Class are further entitled to recover from Defendants the actual or statutory damages Plaintiffs and the Class sustained pursuant to 17 U.S.C. § 1203(c) and for Plaintiffs’ and the Class’s costs and attorneys’ fees in enforcing the Licenses. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of the Licenses.

212. Defendants conspired together and acted jointly and in concert pursuant to their scheme to commit the acts that violated the DMCA alleged herein.

213. Defendants induced Copilot users to unknowingly violate the DMCA by withholding attribution, licensing, and other information as described herein.

---

<sup>40</sup> <https://github.blog/2020-11-19-take-action-dmca-anti-circumvention-and-developer-innovation/#what-dmca-exemptions-do-not-do/>.

<sup>41</sup> <https://github.blog/2021-11-23-copyright-office-expands-security-research-rights/>.

<sup>42</sup> *See, e.g.*, Footnotes 36–41.



**COUNT 2**  
**BREACH OF CONTRACT—OPEN-SOURCE LICENSE VIOLATIONS**  
**California Common Law**  
**(Against All Defendants)**

214. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

215. Plaintiffs and the Class offer code under various Licenses, the most common of which are set forth in Appendix A. Use of each of the Licensed Materials is allowed only pursuant to the terms of the applicable Suggested License.

216. Plaintiffs and the Class granted Defendants a license to copy, distribute, and/or create Derivative Works under the Suggested Licenses. Each of the Suggested Licenses requires at least (1) that attribution be given to the owner of the Licensed Materials used, (2) inclusion of a copyright notice for the Licensed Materials used, and (3) inclusion of the terms of the applicable Suggested License. When providing Output, Copilot does not comply with any of these terms.

217. Defendants accepted the terms of Plaintiffs' and the Class's Licenses when it used the licensed code to create Copilot and when it incorporated the licensed code into Copilot. They have accepted and continue to accept the applicable Licenses every time Copilot Output's Plaintiffs' or the Class's copyrighted code. As such, contracts have been formed between Defendants on the one hand and Plaintiffs and the Class on the other.

218. Plaintiffs and the Class have performed each of the conditions, covenants, and obligations imposed on them by the terms of the License associated with their Licensed Materials.

219. Plaintiffs and members of the Class hold the copyright in the contents of one or more code repositories that have been hosted on GitHub's platform.

220. Plaintiffs and the Class have appended one of the Suggested Licenses to each of the Licensed Materials.

221. Plaintiffs and the Class did not know about, authorize, approve, or license the Defendants' use of the Licensed Materials in the matter at issue in this Complaint before they were used by Defendants.

222. Defendants have substantially and materially breached the applicable Licenses by failing to provide the source code of Copilot nor a written offer to provide the source code upon the request of each licensee.

223. Defendants have substantially and materially breached the applicable Licenses by failing to provide attribution to the creator and/or owner of the Licensed Materials.

224. Defendants have substantially and materially breached the applicable Licenses by failing to include copyright notices when Copilot Outputs copyrighted OS code.

225. Defendants have substantially and materially breached the applicable Licenses by failing to identify the License applicable to the Work and/or including its text when Copilot Outputs code including a portion of a Work.

226. Plaintiffs and the Class have suffered monetary damages as a result of Defendants' conduct.

227. The conduct of Defendants is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

228. As a direct and proximate result of these material breaches by Defendants, Plaintiffs and the Class are entitled to an injunction requiring Defendants to comply with all the terms of any License governing use of code that was used to train Copilot, otherwise incorporated into Copilot, and/or reproduced as Output by Copilot.

229. Plaintiffs and the Class are further entitled to recover from Defendants the damages Plaintiffs and the Class sustained—including consequential damages—for Plaintiffs' and the Class's costs in enforcing their contractual rights. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of contract.

**COUNT 3**  
**BREACH OF CONTRACT — SELLING LICENSED MATERIALS**  
**IN VIOLATION OF GITHUB’S POLICIES**  
**California Common Law**  
**(Against GitHub)**

230. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

231. GitHub’s Privacy Statement, Terms of Service, and GitHub Copilot Terms share definitions and refer to each other. As such, they are collectively referred to herein as “GitHub’s Policies” unless a distinction is necessary and are attached as Exhibit 1.

232. Plaintiffs and the Class are GitHub users who have accepted GitHub’s Policies. As a result, Plaintiffs and the Class have formed a contract with GitHub.

233. Plaintiffs and the Class have performed each of the conditions, covenants, and obligations imposed on them by the terms of GitHub’s Policies.

234. GitHub’s Policies contain multiple explicit provisions that GitHub will not sell the Licensed Materials of the Plaintiffs and Class. GitHub’s Terms of Service document provides that the “License Grant to [GitHub] . . . does not grant GitHub the right to sell Your Content.” Similarly, GitHub’s Privacy Statement defines “personal data” to include “any . . . documents, or other files”, a definition that necessarily comprises source code, and hence the Licensed Materials. (As of May 2023, GitHub has updated this provision on its website to explicitly read “any code, text, . . . documents, or other files”). Elsewhere, the Privacy Statement provides “We do not sell your personal information,” “No selling of personal data,” “We *do not* sell your personal data for monetary or other consideration.” (Emphasis in original).

235. By making the Licensed Materials available through Copilot in violation of the Suggested Licenses, and charging subscription fees, GitHub has been selling Licensed Materials. By selling the Licensed Materials, GitHub has breached these provisions in GitHub’s Policies against selling user data.

236. GitHub has also breached the implied covenant of good faith and fair dealing. GitHub has long held itself out as a good citizen of the global open-source community. GitHub’s Policies were designed to attract Plaintiffs and the Class to become users of the GitHub website



1 by supporting their open-source efforts with fair and ethical practices. By releasing Copilot,  
2 GitHub created a product designed to compete with Plaintiffs and the Class and undermine their  
3 individual open-source communities. In so doing, GitHub did not act fairly or in good faith.

4 237. Plaintiffs and the Class have suffered monetary damages as a result of GitHub's  
5 conduct.

6 238. GitHub's conduct is causing and, unless enjoined and restrained by this Court,  
7 will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be  
8 compensated or measured in money.

9 239. As a direct and proximate result of these material breaches by GitHub, Plaintiffs  
10 and the Class are entitled to an injunction requiring GitHub to comply with all the terms of the  
11 GitHub Policies.

12 240. Plaintiffs and the Class are further entitled to recover from GitHub the damages  
13 Plaintiffs and the Class sustained—including consequential damages—for Plaintiffs' and the  
14 Class's costs in enforcing GitHub's Policies. Plaintiffs and the Class are also entitled to recover as  
15 restitution from GitHub for any unjust enrichment, including gains, profits, and advantages that it  
16 has obtained as a result of its breaches of the GitHub Policies.

17  
18 **COUNT 4**  
19 **INTENTIONAL INTERFERENCE**  
20 **WITH PROSPECTIVE ECONOMIC RELATIONS**  
**California Common Law**  
**(Against All Defendants)**

21 241. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding  
22 and succeeding paragraph as though fully set forth herein.

23 242. Open-source software programmers invite other programmers to view, use and  
24 modify their code and to make changes and improvements to it subject to requirements in certain  
25 licenses. When a programmer uses an open-source software, a contract is formed based on the  
26 terms of the particular open-source license.

27 243. Although no money changes hands in open-source licensing, courts have long  
28 recognized that there are substantial benefits, including economic benefits, to the creation and



distribution of open-source code subject to these open-source licenses. For example, program creators can generate market share for their programs, increase their reputation both nationally and internationally, and discover new improvements to their open-source project. These benefits, however, rely on the proliferation of the licenses and obligations that come with the license along with the code that is subject to that license.

244. GitHub was founded by open-source programmers and has long held itself out as a gathering point for the global open-source community. For open-source programmers, including Plaintiffs and the Class, part of the benefit of becoming a GitHub customer and sharing code there was to make it easier for members of the global open-source community to discover their work, and thereby accrue a user community of contributors and collaborators specific to their open-source projects.

245. User communities create the probability of future economic benefit in a number of ways. Users can provide bug reports, saving authors from having to discover every bug themselves. Users can provide new code that fixes bugs or adds features, saving authors from having to write every line of code themselves. Users sometimes arrange financial contracts with authors for extra licensing rights, or custom features, or technical support. The exposure from a user community can also bring collateral benefits, like job offers or research grants.

246. Plaintiffs and Class members chose to become GitHub customers specifically to avail themselves of these benefits of GitHub and optimize the likelihood of accruing communities of other GitHub customers for their own projects, and to benefit from the future economic benefits likely to arise from those relationships.

247. In other words, Plaintiffs and Class members posted their code on GitHub with an expectation that other programmers would use, modify, copy or otherwise iterate on their posted code subject to the terms of the open-source licenses the code was published subject to.

248. GitHub's status as a focal point of the global open-source community was one of the main reasons Microsoft wanted to own it. At the time it acquired GitHub, Microsoft pledged to uphold these virtues. But Defendants' project of harvesting mass quantities of public open-source code on GitHub for training Codex and Copilot represented an inversion of these

1 priorities. Codex and Copilot essentially act as walled gardens that provide an alternative  
2 interface to the same open-source code, a process sometimes called “vendorization”.

3 249. By failing to provide information about the Suggested Licenses attached to the  
4 Licensed Materials, Defendants intentionally prevented Copilot users from becoming part of the  
5 user communities that would ordinarily accrete around the open-source projects of Plaintiffs and  
6 the Class. Instead, Defendants reserved those benefits for themselves.

7 250. Defendants knew that they were interfering with Plaintiffs and Class members’  
8 prospective open-source relationships because Defendants knew that Codex and Copilot were  
9 emitting code subject to open-source licenses without the licenses attached.

10 251. Defendants have therefore intentionally and wrongfully interfered with the  
11 prospective business interests and expectations of Plaintiffs and the Class.

12 252. Plaintiffs have been deprived of the economic benefits of open-source licenses.  
13 Plaintiffs and the Class have suffered monetary, reputational, and other damages as a result of  
14 Defendants’ conduct.

15 253. Unless enjoined and restrained by this Court, Defendants’ conduct will continue  
16 to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or  
17 measured in money.

18  
19 **COUNT 5**  
20 **NEGLIGENT INTERFERENCE**  
21 **WITH PROSPECTIVE ECONOMIC RELATIONS**  
**California Common Law**  
**(Against All Defendants)**

22 254. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding  
23 and succeeding paragraph as though fully set forth herein.

24 255. As described in Paragraphs 242 through 247 herein, Plaintiffs and Class members  
25 post their code subject to open-source licenses in order to enjoy the economic benefits associated  
26 with the distribution of open-source software code.

27 256. Defendants knew or should have known that Plaintiffs and the Class chose to  
28 become GitHub customers specifically to optimize the likelihood of accruing communities of

1 other GitHub customers for their own projects, and to benefit from the future economic benefits  
2 likely to arise from those open-source licensing relationships.

3 257. Defendants knew or should have known that by scraping the Licensed Materials of  
4 Plaintiffs and the Class and using it to create competing products (namely, Codex and Copilot)  
5 that did not adhere to the obligations of the Suggested Licenses, that Defendants would disrupt  
6 the formation and growth of these user communities, and also disrupt the economic benefits that  
7 would ordinarily accrue to Plaintiffs and the Class from the growth of those user communities.

8 258. Defendants' conduct falls far outside the boundaries of fair competition because  
9 Defendants have leveraged GitHub's status as a gathering point for the global open-source  
10 community to undermine that very community, including Plaintiffs and the Class. Defendants  
11 have leveraged the permissive nature of the Suggested Licenses to undermine the Licensed  
12 Materials, thereby harming Plaintiffs and the Class.

13 259. Defendants' conduct could not have been performed without a direct effect on  
14 Plaintiffs' economic interests as Plaintiffs' open-source code was hosted on GitHub and used to  
15 train Codex and Copilot.

16 260. Plaintiffs uploaded their open-source software code on GitHub with the  
17 reasonable expectation that other open-source programmers would use, modify, copy or  
18 otherwise iterate on their code.

19 261. The adverse effect on Plaintiffs was foreseeable as Defendants are aware of the  
20 obligations that carry with open-source licenses. Further, Defendants knew that Codex and  
21 Copilot copied code used for training the models without the associated licenses attached and  
22 without any of the necessary obligations that carry with the licenses, e.g., attribution.

23 262. Even after Defendants knew that Codex and Copilot would reproduce code  
24 verbatim, including in some instances, with the incorrect licenses or license text, Defendants  
25 continued to operate Codex and Copilot. In other words, Defendants continued to operate Codex  
26 and Copilot after they should have known that Codex and Copilot's operation were depriving  
27 Plaintiffs of the economic benefits of open-source distribution.



263. Defendants have therefore negligently and wrongfully interfered with the prospective business interests and expectations of Plaintiffs and the Class.

264. Plaintiffs and the Class have suffered monetary, reputational, and other damages as a result of Defendants' conduct.

265. Unless enjoined and restrained by this Court, Defendants' conduct will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

**COUNT 6**  
**UNJUST ENRICHMENT**  
**California Common Law**  
**(Against All Defendants)**

266. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

267. Plaintiffs and the Class have invested substantial time and energy in creating the Licensed Materials.

268. Plaintiffs hosted their Licensed Materials on GitHub subject to open-source licenses.

269. Defendants have unjustly utilized access to Licensed Materials hosted on GitHub to create Codex and Copilot. Specifically, Defendants used Plaintiffs' Licensed Materials to train Codex and Copilot without following the licenses under which the Licensed Materials were published.

270. By using Plaintiffs' Licensed Materials to train Codex and Copilot, Plaintiffs and the Class were deprived of the benefits of their open-source licenses, including monetary damages.

271. Plaintiffs did not consent to the unauthorized use of their Licensed Materials to train Codex and Copilot.

272. Defendants derived profit or other benefits from the use of the Licensed Materials to train Codex and Copilot.

273. It would be unjust for Defendants to retain those benefits.



274. The conduct of Defendants is causing and, unless enjoined and restrained by this Court, will continue to cause Plaintiffs and the Class great and irreparable injury that cannot fully be compensated or measured in money.

**COUNT 7**  
**UNFAIR COMPETITION**  
**Cal. Bus. & Prof. Code §§ 17200, *et seq.*; and California Common Law**  
**(Against All Defendants)**

275. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

276. Defendants have engaged in unlawful business practices, including:

- a. Violations of Plaintiffs' and the Class's rights under the DMCA;
- b. Violations of Plaintiffs' and the Class's rights under the open-source licenses their Licensed Materials was published subject to;
- c. Tortious interference in Plaintiffs' and the Class's prospective economic advantage with users of their code;
- d. Failing to attribute Codex and Copilot's Output as originating Plaintiffs and the Class rather than from Copilot, GitHub, and/or OpenAI;

The details of the unlawful business practices are set forth herein.

277. The unlawful business practices described herein violate California Business and Professions Code section 17200 *et seq.* because that conduct is otherwise unlawful including under the DMCA.

278. The unlawful business practices described herein likewise violate California Business and Professions Code section 17200 *et seq.* because it is unfair because they are immoral, unethical, oppressive, unscrupulous or injurious to consumers because, among other reasons, Defendants pass off Codex and Copilot's output without proper attribution as required by the licenses Plaintiffs' and Class members' code was published subject to.

279. The unlawful business practices described herein likewise violate California Business and Professions Code section 17200 *et seq.* as fraudulent because consumers are likely to be deceived because, among other reasons, Defendants cause Codex and Copilot's output to be

emitted without the proper licensing and attribution required by the licenses Plaintiffs and Class members publish their material subject to.

280. The unlawful business practices described herein violate the common law because Plaintiffs and Class members invested substantial time, skill and money in developing the Licensed Materials. Defendants misappropriated and used Plaintiffs and Class members' Licensed materials without authorization or consent in order to, *inter alia*, train and develop Codex and Copilot.

281. Plaintiffs and the Class have suffered economic injury as a result of Defendants' conduct. Specifically, there are economic benefits to the creation of open-source works such as generating market share for programs, increasing national or international reputation by incubating open-source projects, and deriving value from improvements to software based on suggestions by end-users.

**COUNT 8  
NEGLIGENCE  
California Common Law  
(Against All Defendants)**

282. Plaintiffs and the Class hereby repeat and incorporate by reference each preceding and succeeding paragraph as though fully set forth herein.

283. Defendants owed a duty of reasonable care toward Plaintiffs and the Class based upon Defendants' relationship to them. This duty is based upon Defendants' contractual obligations, custom and practice, right to control information in its possession, exercise of control over the information in its possession, authority to control the information in its possession, and the commission of affirmative acts that resulted in said harms and losses. Additionally, this duty is based on the requirements of California Civil Code section 1714 requiring all "persons," including Defendants, to act in a reasonable manner toward others.

284. Defendants breached their duties by negligently, carelessly, and recklessly collecting, maintaining, and controlling their customers' Licensed Materials and engineering, designing, maintaining, and controlling systems—including Codex and Copilot—which are trained on Plaintiffs' and Class members' Licensed Materials without their authorization.

285. Microsoft and GitHub owed its users a duty of care to safeguard and maintain Licensed Materials on its website and to prevent unauthorized use of the Licensed Materials.

286. Microsoft and GitHub also owed its user a duty of care not to itself use the Licensed Materials in a way that would foreseeably cause Plaintiffs and Class members injury, for instance, by using Licensed Materials to train Copilot.

287. OpenAI owed Plaintiffs and Class members a duty of care by using open-source code in violation of open-source licenses to train Codex and Copilot.

288. Defendants, through their unlawful acts described herein, breach of their duties proximately caused Plaintiffs and Class members injuries.

289. Plaintiffs and Class members' injuries were foreseeable because Defendants are aware of the benefits of open-source licensing because they hold themselves out as advocates for open source, and serve as platforms to host open-source software.

## **IX. DEMAND FOR JUDGMENT**

**WHEREFORE**, Plaintiffs requests that the Court enter judgment on their behalf and on behalf of the Class defined herein, by adjudging and decreeing that:

290. This action may proceed as a class action, with Plaintiffs serving as Class Representatives, and with Plaintiffs' counsel as Class Counsel;

- a) Judgment in favor of Plaintiffs and the Class and against Defendants;
- b) Permanent injunctive relief, including but not limited to making changes to its Copilot product to ensure that all applicable information set forth in 17 U.S.C. § 1203(b)(1) is included in along with any Output including associated code;
- c) An order of costs and allowable attorney's fees pursuant to 17 U.S.C. § 1203(b)(4)–(5);
- d) An award of statutory damages pursuant to 17 U.S.C. § 1203(b)(3) and 17 U.S.C. §



1203(c)(3),<sup>43</sup> or, in the alternative, an award of actual damages and any additional profits pursuant to 17 U.S.C. § 1203(c)(2) (including tripling damages pursuant to 17 U.S.C. § 1203(c)(4) if applicable);

- e) An award of damages for harms resulting from Defendants’ breach of Licenses;
- f) An award of damages, including punitive damages, for harms resulting from Defendants’ tortious interference in Plaintiffs’ and the Class’s prospective economic advantage;
- g) An award of damages in the amount Defendants have been unjustly enriched through their conduct as alleged herein as well as punitive damages in connection with this conduct;
- h) An award of damages, including punitive damages, for harms resulting from Defendants’ acts of unfair competition;
- i) An award of damages for harms resulting from GitHub’s breach of the GitHub Policies; and
- j) An award of damages, including punitive damages, for harms resulting from Defendants’ negligence including in the failure to control Plaintiffs’ and the Class’s Licensed Materials.

291. Injunctive relief sufficient to alleviate and stop Defendants’ unlawful conduct alleged herein.

292. Plaintiffs and the Class are entitled to prejudgment and post-judgment interest on the damages awarded them, and that such interest be awarded at the highest legal rate from and after the date this class action complaint is first served on Defendants;

---

<sup>43</sup> Plaintiffs estimate that statutory damages for Defendants’ direct violations of DMCA Section 1202 alone will exceed \$9,000,000,000. That figure represents minimum statutory damages (\$2,500) incurred three times for each of the 1.2 million Copilot users Microsoft reported in June 2022. Each time Copilot provides an unlawful Output it violates Section 1202 three times (distributing the Licensed Materials without: (1) attribution, (2) copyright notice, and (3) License Terms). So, if each user receives just one Output that violates Section 1202 throughout their time using Copilot (up to fifteen months for the earliest adopters), then GitHub and OpenAI have violated the DMCA 3,600,000 times. At minimum statutory damages of \$2500 per violation, that translates to \$9,000,000,000.



293. Defendants are to be jointly and severally responsible financially for the costs and expenses of a Court approved notice program through post and media designed to give immediate notification to the Class.

294. Plaintiffs and the Class receive such other or further relief as may be just and proper.

## X. JURY TRIAL DEMANDED

Pursuant to Federal Rule of Civil Procedure 38(b), Plaintiffs demand a trial by jury of all the claims asserted in this Complaint so triable.

Dated: June 8, 2023

By: /s/ Joseph R. Saveri  
Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)  
Steven N. Williams (State Bar No. 175489)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Louis A. Kessler (State Bar No. 243703)  
Elissa A. Buchanan (State Bar No. 249996)  
Travis Manfredi (State Bar No. 281779)  
**JOSEPH SAVERI LAW FIRM, LLP**  
601 California Street, Suite 1000  
San Francisco, California 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
Email: jsaveri@saverilawfirm.com  
swilliams@saverilawfirm.com  
czirpoli@saverilawfirm.com  
cyoung@saverilawfirm.com  
lkessler@saverilawfirm.com  
eabuchanan@saverilawfirm.com  
tmanfredi@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)  
1920 Hillhurst Avenue, #406  
Los Angeles, CA 90027  
Telephone: (323) 968-2632  
Facsimile: (415) 395-9940  
Email: mb@buttericklaw.com  
*Counsel for Plaintiffs and the Proposed Class*

# EXHIBIT C

Joseph R. Saveri (State Bar No. 130064)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Louis A. Kessler (State Bar No. 243703)  
Elissa A. Buchanan (State Bar No. 249996)  
William W. Castillo Guardado (State Bar No. 294159)  
Holden Benon (State Bar No. 325847)  
**JOSEPH SAVERI LAW FIRM, LLP**  
601 California Street, Suite 1000  
San Francisco, California 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
Email: jsaveri@saverilawfirm.com  
czirpoli@saverilawfirm.com  
cyoung@saverilawfirm.com  
lkessler@saverilawfirm.com  
eabuchanan@saverilawfirm.com  
wcastillo@saverilawfirm.com  
hbenon@saverilawfirm.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
OAKLAND DIVISION**

J. DOE 1, et al.,  
Individual and Representative Plaintiffs,  
v.  
GITHUB, INC., et al.,  
Defendants.

Case Nos. 4:22-cv-06823-JST  
4:22-cv-07074-JST

**PLAINTIFFS' MOTION FOR  
LEAVE TO FILE MOTION FOR  
RECONSIDERATION (L.R. 7-9)**

**NOTICE OF MOTION AND  
MOTION FOR LEAVE TO FILE MOTION FOR RECONSIDERATION**

**TO DEFENDANTS AND THEIR COUNSEL OF RECORD:**

Please take notice that pursuant to Civil Local Rule 7-9, Plaintiffs Doe 1, Doe 2, Doe 3, Doe 4, and Doe 5 (“Plaintiffs”) hereby move this Court for leave to file a motion for reconsideration of this Court’s January 3, 2024 Order Granting In Part Denying In Part Motion to Dismiss the First Amended Complaint (ECF No. 189) (hereinafter “Order”).

The requirements of Local Rule 7-9(b)(1) and (b)(3) are met here because a material difference in law exists from that which was presented to the Court before entry of the Order, and that in entering the Order, the Court failed to consider material facts and dispositive legal arguments which were presented to the Court before the Order.

**MEMORANDUM OF POINTS AND AUTHORITIES**

Plaintiffs seek leave to move for reconsideration of the Court’s dismissal of Plaintiffs’ damages and injunctive relief claims under the Digital Millennium Copyright Act (“DMCA”) in the Order. ECF No. 189.

**I. STATEMENT OF RELEVANT FACTS**

On November 3, 2022, Plaintiffs filed their initial complaint (ECF No. 1, “Initial Complaint”) alleging, among other things, that Defendants violated DMCA §§ 1202(b)(1) and (b)(3) by removing or altering Copyright Management Information (“CMI”) from Plaintiffs’ Licensed Materials and distributing copies of those materials that did not include CMI. Initial Complaint, ¶¶ 138–167. The five named Plaintiffs—Doe 1, Doe 2, Doe 3, Doe 4 and Doe 5—on behalf of a class of similarly situated plaintiffs, alleged claims for injunctive relief and damages under federal statutes, state statutes and the common law. Initial Complaint, Counts I–XII, ECF No. 1.

In their initial motions to dismiss, Defendants argued that DMCA §§ 1202(b)(1) and (b)(3) require that the copy that had its CMI removed or altered and/or that was distributed without CMI must be an identical copy of the work. ECF Nos. 50 at 13 (Microsoft and GitHub Motion to Dismiss Complaint); and 53 at 9-10 (OpenAI Motion to Dismiss Complaint). In opposition,



1 Plaintiffs argued that DMCA §§ 1202(b)(1) and (b)(3) contain no such requirement and  
2 distinguished the cases cited by Defendants. ECF Nos. 66 at 7–8 (Opposition to GitHub and  
3 Microsoft MTD Complaint); and 67 at 17–19 (Opp. to OpenAI MTD Complaint).

4 After a hearing on the motions, the Court denied Defendants’ motions as to DMCA  
5 §§ 1202(b)(1) and 1202(b)(3). The Court noted that “Plaintiffs thus plead sufficient facts to  
6 support a reasonable inference that Defendants intentionally designed the programs to remove CMI  
7 from any licensed code they reproduce as output.” ECF No. 95 at 19 (May 11, 2023 order). This  
8 Court ruled that Plaintiffs had pleaded facts establishing standing to pursue a claim for injunctive  
9 relief under DMCA §§ 1202(b)(1) and (b)(3) but had failed to plead facts to pursue a claim for  
10 damages. ECF No. 95 at 10 (“Plaintiffs ... have standing to pursue injunctive relief”) and 8  
11 (“Because Plaintiffs do not allege that they themselves have suffered the injury they describe, they  
12 do not have standing to seek retrospective relief for that injury”).

13 On June 8, 2023, Plaintiffs filed their First Amended Complaint (“FAC”), seeking both  
14 damages and injunctive relief as part of their DMCA claim. ECF No. 97; 102 (corrected FAC); 135  
15 (second corrected FAC). The FAC maintained the allegations the Court found sufficient to  
16 establish standing for injunctive-relief claims. In order to address the deficiencies with respect to  
17 standing to pursue damages claims, the FAC included new facts with respect to Doe 1, Doe 2, and  
18 Doe 5. With respect to each, Plaintiffs offered specific examples of the code written by each as well  
19 as the output from Copilot. FAC ¶¶ 97-128. Plaintiffs alleged, based on these facts, that the output  
20 from Copilot is often a verbatim copy and often a modification. *Id.* Plaintiffs alleged these facts  
21 were sufficient to state a claim by Doe 1, Doe 2, and Doe 5, on their own behalf and on behalf of the  
22 class they seek to represent. The allegations with respect to Doe 3 and Doe 4 were undisturbed,  
23 thus leaving unaffected the allegations the Court found sufficient for injunctive relief as to them.  
24 More generally, the FAC did not change the allegations the Court found sufficient to establish  
25 standing for injunctive relief as to any of the named Plaintiffs.

26 With respect to damages, Plaintiffs alleged that Copilot reproduced copies of Doe 5’s code  
27 verbatim. FAC ¶ 123 (“The first suggestion from Copilot offers . . . a verbatim copy of Doe 5’s  
28 original code”) and ¶ 125 (“Once again, the first suggestion from Copilot offers . . . a verbatim copy

of Doe 5’s code (except for small cosmetic variations in line breaks)). Defendants moved only to dismiss Plaintiffs’ damage claims (and not their injunctive-relief claims) and repeated their contention that claims under DMCA §§ 1202(b)(1) and (b)(3) require removal or alteration of CMI from an identical copy of a copyrighted work. ECF No. 108 at 13–15 (Microsoft and GitHub MTD FAC); and 109 at 12–14 (OpenAI MTD FAC).

In opposition, Plaintiffs highlighted that “this argument has been extensively (and exhaustively) briefed,” and that the Court had already considered the arguments raised in the prior briefing. ECF Nos. 140 at 13-14; 141 at 14.

In its January 3, 2024 Order (ECF No. 189), the Court granted the Defendants’ motions as to DMCA §§ 1202(b)(1) and 1202(b)(3), in their entirety. The Court granted the motion with respect to both injunctive relief and damages, including the injunctive relief claims previously found sufficient and including the damage claims of Doe 5. *See* ECF No. 95. The Court found that Plaintiffs failed to allege sufficient facts of identity of output. The Court did not acknowledge the allegations that CoPilot emits identical verbatim copies of Doe5’s code.

## II. LEGAL ARGUMENT

First, Plaintiffs did not change the allegations with respect to two of the Plaintiffs, Doe 3 and Doe 4, in any way and did not change the allegations found sufficient for injunctive relief as to Doe 1, Doe 2 and Doe 5. Therefore, the facts that the Court found sufficient for injunctive relief remained and would logically continue to be sufficient in the FAC and subsequent amendments.

Second, for that reason, and for the additional reason that Defendants did not challenge them, Plaintiffs did not reiterate the arguments which the Court had previously found persuasive with respect to injunctive relief standing. Plaintiffs understood the issue to have been resolved in their favor in the Court’s prior order on those motions. *See* ECF No. 95. Repeating arguments which the Court found persuasive, and which the Defendants did not repeat, was unnecessary. The Court should not have dismissed them.

Third, the Court failed to consider case law which holds that DMCA § 1202(b) applies, contrary to Defendants’ arguments, to non-identical collaborative or derivative works. The Court

also disregarded well-pleaded allegations with respect to DOE 5 that Copilot emits verbatim copies of Doe 5's original code. *See* FAC ¶ 123, 125.

**A. By the Court's reasoning in its Order, the DMCA injunctive claims for Does 3 and 4 should have remained intact, but instead they were dismissed.**

In its Order, the Court analyzed standing with respect to each plaintiff. The Court stated at the outset that "Does 1, 2, and 5 have standing to pursue claims for both injunctive relief and damages, whereas Does 3 and 4 have standing to pursue only claims for injunctive relief." Order at 9.

Later in the Order, the Court acknowledged that Plaintiffs' FAC makes allegations of identical output from Copilot that are not mutually exclusive with allegations of modified output: "Plaintiffs' amended complaint alleges that '[t]hough Output from Copilot is often a verbatim copy, even more often it is a modification." Order at 15.

Based on code examples added to the FAC for Does 1, 2, and 5, the Court ruled that because "the examples Plaintiffs provide with respect to Does 1, 2, and 5 state that the Copilot output is a 'modified format' ... [t]his ... is not sufficient for a Section 1202(b) claim." Order at 15 (emphasis added). But Plaintiffs did not make equivalent allegations of code examples for Does 3 and 4, as the Court acknowledged by not mentioning them.

Despite this salient difference in the allegations for Does 3 and 4, at the end of the Order, the Court ruled that "Defendants' motions to dismiss Plaintiffs' claims under Sections 1202(b)(1) and 1202(b)(3) of the DMCA ... are granted." Order at 16. That is, the Court dismissed Plaintiffs' DMCA claims in their entirety—as to both damages and injunctive relief—despite the differences in the unchanged allegations for Does 3 and 4 and despite the fact that the Court had previously found them sufficient for purposes of standing to pursue injunctive relief claims.

**B. The Court failed to consider caselaw holding that identical copies are not required to bring a DMCA claim.**

With respect to standing for the damages claims by Doe 5, Plaintiffs also move for reconsideration. Plaintiffs understand that "[m]otions for reconsideration are designed to bring to the Court's attention clear instances of missed arguments, not simply to make the very same points



but more loudly.” *Sheet Metal Workers Nat’l Pension Fund v. Bayer Aktiengesellschaft*, No. 20-CV-04737-RS, 2021 WL 5302525, at \*1 (N.D. Cal. Nov. 15, 2021).

In dismissing Plaintiffs’ DMCA claim for damages, the Court relied on three district court cases, *Advanta-STAR Auto. Rsch. Corp. of Am. v. Search Optics, LLC*, No. 22-CV-1186 TWR (BLM), 2023 WL 3366534 (S.D. Cal. May 9, 2023), *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG, 2020 WL 5991503 (C.D. Cal. Aug. 14, 2020), and *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. CIV. 13-00496 SOM, 2015 WL 263556 (D. Haw. Jan. 21, 2015), *aff’d*, 700 F. App’x 674 (9th Cir. 2017). Order at 15. In so doing, the Court failed to consider the decision in *GC2 v. Int’l Game Tech.*, 391 F. Supp. 3d 828 (N.D. Ill. 2019), despite its inclusion in Plaintiffs’ oppositions to Defendants’ first motions to dismiss. *See* ECF Nos. 66 & 67. As Plaintiffs argued, the Court in *GC2* held that DMCA § 1202(b) applies to collaborative or derivative works that are not identical, noting that:

[None of the] cases cited by the defendants stands for the broad proposition that derivative or collaborative works are categorically excluded from protection under the DMCA’s provision for removal of copyright management information. Indeed, the “original work” language on which the defendants precariously rest their entire argument does not even appear in the statute. Rather, it is entirely a product of district court opinions focused on different language in provision.

391 F. Supp. 3d at 843–44; ECF No. 67 at 18. Notably, neither the Court’s May 11, 2023 Order nor its January 3, 2024 Order apply the holding of *GC2* to the facts here or otherwise reference the case. *See* ECF Nos. 95 and 189.<sup>1</sup> The decision in *GC2*, unlike those cited by the Court in its January 3, 2024 Order, is directly on point and provides a sound basis for the correct statutory analysis of DMCA § 1202(b).

<sup>1</sup> Plaintiffs note the decision in *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 429 (S.D. Tex. 2023). Applying *GC2*, the Court held that “none of the cases Defendants cite support the proposition that Section 1202 requires a plaintiff to plead the allegedly infringing works are identical copies of the plaintiff’s works.” 667 F. Supp. 3d at 429. Plaintiffs did not cite the decision in any of the prior briefing on the issue of identity.



**C. Though the Court’s Order initially acknowledged allegations in Plaintiffs’ FAC regarding verbatim output of Doe 5’s code, the Court did not account for these allegations in its ruling.**

In addition, the Court overlooked where Plaintiffs specifically alleged identical output of Doe 5’s Licensed Materials. Specifically, Plaintiffs alleged, “Once again, the first suggestion from Copilot offers ... a verbatim copy of Doe 5’s code (except for small cosmetic variations in line breaks).” FAC ¶ 125. The Court even acknowledged another instance of identical output from Doe 5, quoting the FAC’s allegation that “Copilot offers to complete the second test with a verbatim copy of Doe 5’s original code.” Order at 4 (citing to FAC ¶ 122-23).

The Court nevertheless found that Plaintiffs had failed to plead identity of Output to support their Section 1202(b) claims. *See* ECF No. 189 at 2, 14–16 (“Indeed, the examples Plaintiffs provide with respect to Does 1, 2, and 5 state that the Copilot output is a “modified format,” “variation[,]” or the “functional[] equivalent” of the licensed code”). This is inconsistent with the well-plead allegations of the complaint. In doing so, the Court failed to construe the allegations in the FAC as true and failed to construe the pleadings in the light most favorable to Plaintiffs. *See Knievel v. ESPN*, 393 F.3d 1068, 1072 (9th Cir. 2005). The Court focused solely on Plaintiffs’ examples of near-identical Outputs and failed to consider Plaintiffs’ allegations of identical outputs. Thus, even if the Court maintains its ruling that claims under DMCA § 1202 require identity, Doe 5 met this burden in the FAC and the Court’s dismissal of Doe 5’s claim for damages should be revoked and their claim reinstated.

**III. CONCLUSION**

As the Court previously held, the allegations with respect to Doe 3 and Doe 4 were sufficient to establish standing injunctive relief. Because these allegations were not changed in the First Amended Complaint, they should not have been dismissed. With respect to Doe 5 Plaintiffs’ claims under DMCA §§ 1202(b)(1) and (b)(3) were improperly dismissed by this Court. Plaintiffs’ claims should not have been dismissed because identity is not required for claims under DMCA §§ 1202(b)(1) and (b)(3) under existing authority. Even if identity is required, Plaintiffs’ FAC set forth allegations of identical (“verbatim”) output of Doe 5’s code.

For these reasons, this Court should amend its Order to specify the injunctive relief claims are not dismissed and to deny Defendants' motions to dismiss Plaintiffs' claims for damages as to Doe 5 under DMCA §§ 1202(b)(1) and 1202(b)(3).

Dated: February 28, 2024

By: /s/ Joseph R. Saveri  
Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Louis A. Kessler (State Bar No. 243703)  
Elissa A. Buchanan (State Bar No. 249996)  
William W. Castillo Guardado (State Bar No. 294159)  
Holden Benon (State Bar No. 325847)  
JOSEPH SAVERI LAW FIRM, LLP  
601 California Street, Suite 1000  
San Francisco, California 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
Email: jsaveri@saverilawfirm.com  
czirpoli@saverilawfirm.com  
cyoung@saverilawfirm.com  
lkessler@saverilawfirm.com  
eabuchanan@saverilawfirm.com  
wcastillo@saverilawfirm.com  
hbenon@saverilawfirm.com

*Counsel for Plaintiffs and the Proposed Class*

# EXHIBIT D



Joseph R. Saveri (State Bar No. 130064)  
Cadio Zirpoli (State Bar No. 179108)  
Christopher K.L. Young (State Bar No. 318371)  
Louis A. Kessler (State Bar No. 243703)  
Elissa A. Buchanan (State Bar No. 249996)  
William W. Castillo Guardado (State Bar No. 294159)  
Holden J. Benon (State Bar No. 325847)

**JOSEPH SAVERI LAW FIRM, LLP**

601 California Street, Suite 1000  
San Francisco, CA 94108  
Telephone: (415) 500-6800  
Facsimile: (415) 395-9940  
Email: jsaveri@saverilawfirm.com  
czirpoli@saverilawfirm.com  
cyoung@saverilawfirm.com  
lkessler@saverilawfirm.com  
eabuchanan@saverilawfirm.com  
wcastillo@saverilawfirm.com  
hbenon@saverilawfirm.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*

[Additional Counsel Listed on Signature Page]

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
OAKLAND DIVISION**

J. DOE 1 et al.,  
*Individual and Representative Plaintiffs,*  
v.  
GITHUB, INC., et al.,  
*Defendants.*

Case Nos. 4:22-cv-06823-JST  
4:22-cv-07074-JST

**PLAINTIFFS' OPPOSITION TO  
OPENAI'S MOTION TO DISMISS THE  
SECOND AMENDED COMPLAINT**

Date: May 16, 2024  
Time: 2:00 p.m.  
Courtroom: 6, 2nd Floor  
Judge: Hon. Jon Tigar

## TABLE OF CONTENTS

I.	INTRODUCTION .....	1
II.	PROCEDURAL HISTORY .....	1
A.	Plaintiffs’ Initial Complaint (ECF No. 1) .....	1
B.	The Court’s Order on OpenAI’s Motion to Dismiss the Initial Complaint (ECF No. 95) .....	2
C.	Plaintiffs’ First Amended Complaint (ECF No. 97) .....	2
D.	The Court’s Order on OpenAI’s Motion to Dismiss First Amended Complaint (ECF No. 189).....	3
E.	Plaintiffs’ SAC (ECF No. 200) and Its New Allegations .....	3
III.	ARGUMENT .....	3
A.	Plaintiffs’ Second Amended Complaint Plausibly Alleges DMCA Violations under DMCA §§ 1202(b)(1) and 1202(b)(3) .....	3
1.	Plaintiffs’ DMCA claims are plausibly alleged under any standard.....	5
a.	Identity is not an element of a DMCA § 1202 claim .....	6
b.	Plaintiffs allege that Copilot reproduces identical copies of Plaintiffs’ code .....	10
2.	The DMCA does not require an entire work to be copied.....	12
B.	Plaintiffs Adequately Allege Breach of Contract by OpenAI Through Its Mass Violation of Open-Source Licenses.....	13
1.	OpenAI’s failure to raise these arguments before forecloses OpenAI’s Rule 12 motion under Rule 12(g)(2) .....	13
2.	Plaintiffs adequately allege breach of contract based on the operation of OpenAI’s Codex product and its integration into GitHub and OpenAI’s joint venture, the Copilot product .....	15
3.	OpenAI’s failure to abide by open-source licenses adequately supports a breach-of-contract claim .....	17
a.	Plaintiffs sufficiently plead breach of contract because the attribution and notice requirements of OSLs are properly considered covenants and not conditions .....	18

1	a.	Jacobsen v. Katzer is not determinative.....	20
2	b.	Plaintiffs’ allegations of OSL violations include “copyleft”	
3		requirements that OpenAI completely fails to address.....	24
4	IV.	CONCLUSION .....	25



## TABLE OF AUTHORITIES

### Cases

<i>ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.</i> ,	
667 F. Supp. 3d 411 (S.D. Tex. 2023) .....	6, 7, 8
<i>In re Apple iPhone Antitrust Litig.</i> ,	
846 F.3d 313 (9th Cir. 2017).....	13
<i>Artifex Software, Inc. v. Hancom, Inc.</i> ,	
No. 16-CV-06982-JSC, 2017 WL 1477373 (N.D. Cal. Apr. 25, 2017) .....	17, 24, 25
<i>Burroughs v. Metro-Goldwyn-Mayer, Inc.</i> ,	
683 F.2d 610 (2d Cir. 1982) .....	12
<i>Call v. Alcan Pac. Co.</i> ,	
251 Cal.App.2d 442 (1967).....	19
<i>Columbia Pictures Indus., Inc. v. Galindo</i> ,	
No. 20-cv-3129-MEMF, 2022 WL 17094713 (C.D. Cal. Nov. 18, 2022) .....	6
<i>Doe 1 v. GitHub, Inc.</i> ,	
672 F. Supp. 3d 837 (N.D. Cal. 2023).....	2, 5, 13, 17
<i>Fischer v. Forrest</i> ,	
286 F. Supp. 3d 590 (S.D.N.Y. 2018) .....	7, 8
<i>Forest v. Equitable Life Assurance Soc’y of U.S.</i> ,	
No. C99-5173 SI, 2001 WL 1338809 (N.D. Cal. June 12, 2001) .....	16
<i>Friedman v. Live Nation Merch., Inc.</i> ,	
833 F.3d 1180 (9th Cir. 2016).....	8
<i>Frost-Tsuji Architects v. Highway Inn, Inc.</i> ,	
No. CIV 13-00496 SOM, 2015 WL 263556 (D. Haw. Jan. 21, 2015) .....	7
<i>Harrell v. City of Gilroy</i> ,	
No. 17-CV-05204-LHK, 2019 WL 452039 (N.D. Cal. Feb. 5, 2019) .....	14
<i>Harrington v. Pinterest, Inc.</i> ,	
No. 5:20-CV-05290-EJD, 2022 WL 4348460 (N.D. Cal. Sept. 19, 2022) .....	4
<i>Horgan v. Macmillan, Inc.</i> ,	
789 F.2d 157 (2d Cir. 1986).....	12
<i>Jacobsen v. Katzer</i> ,	
535 F.3d 1373 (Fed. Cir. 2008).....	<i>passim</i>

1	<i>Kelly v. Arriba Soft Corp.</i> ,	
2	77 F. Supp. 2d 1116 (C.D. Cal. 1999) .....	7
3	<i>Kirk Kara Corp. v. W. Stone &amp; Metal Corp.</i> ,	
4	No. CV 20-1931-DMG (EX), 2020 WL 5991503 (C.D. Cal. Aug. 14, 2020) .....	3, 7, 8
5	<i>Laborers Health Welf. Tr. Fund v. Kaufman &amp; Broad of N. Cal., Inc.</i> ,	
6	707 F.2d 412 (9th Cir. 1983) .....	20
7	<i>Logan v. Meta Platforms, Inc.</i> ,	
8	636 F. Supp. 3d 1052, 2022 WL 14813836 (N.D. Cal. Oct. 25, 2022) .....	4
9	<i>Mario V. v. Armenta</i> ,	
10	No. 18-cv-00041-BLF, 2019 WL 81337140 (N.D. Cal. Apr. 17, 2019) .....	14
11	<i>MDY Indus., LLC v. Blizzard Ent., Inc.</i> ,	
12	629 F.3d 928 (9th Cir. 2010) .....	18
13	<i>Mollman v. Zoetop Bus. Co.</i> ,	
14	No. CV 22-4128 (PA), 2022 WL 17207103 (C.D. Cal. Sept. 16, 2022) .....	4
15	<i>Murphy v. Millennium Radio Grp. LLC</i> ,	
16	650 F.3d 295 (3d Cir. 2011) .....	8
17	<i>Myrick v. Mastagni</i> ,	
18	185 Cal. App. 4th 1082 (2010) .....	16
19	<i>Netbula, LLC v. Storage Tech. Corp.</i> ,	
20	No. C06-07391 MJJ, 2008 WL 228036 (N.D. Cal. Jan. 18, 2008) .....	19, 24
21	<i>Oracle Int’l Corp. v. Rimini St., Inc.</i> ,	
22	No. 2:14-cv-01699-MMD-DJA, 2023 WL 4706127 (D. Nev. July 24, 2023) .....	9, 11, 12
23	<i>Pac. Allied v. Century Steel Prods.</i> ,	
24	162 Cal. App. 2d 70, 327 P.2d 547 (1958) .....	19
25	<i>In re Packaged Seafood Prods. Antitrust Litig.</i> ,	
26	277 F. Supp. 3d 1167 (S.D. Cal. 2017) .....	14, 15
27	<i>ProCD, Inc. v. Zeidenberg</i> ,	
28	86 F.3d 1447 (7th Cir. 1996) .....	21, 22
	<i>Shaw v. Lindheim</i> ,	
	919 F.2d 1353 (9th Cir. 1990) .....	12
	<i>Software Freedom Conservancy, Inc. v. Vizio, Inc.</i> ,	
	No. 8:21-cv-01943-JLS-KES, 2022 WL 1527518 (C.D. Cal. May 13, 2022) .....	17, 22, 25

1	<i>Software Pricing Partners, LLC v. Geisman,</i>	
2	No. 3:19-cv-00195-RJC-DCK, 2022 WL 3971292 (W.D.N.C. Aug. 31, 2022).....	9
3	<i>Splunk Inc. v. Cribl, Inc.,</i>	
4	662 F. Supp. 3d 1029 (N.D. Cal. 2023).....	9
5	<i>Stevens v. Corelogic, Inc.,</i>	
6	899 F.3d 666 (9th Cir. 2018).....	3
7	<i>Textile Secrets Int’l, Inc. v. Ya-Ya Brand Inc.,</i>	
8	524 F. Supp. 2d 1184 (C.D. Cal. 2007).....	6
9	<i>Ticketmaster L.L.C. v. Prestige Ent., Inc.,</i>	
10	306 F. Supp. 3d 1164 (C.D. Cal. 2018).....	<i>passim</i>
11	<i>Versata Software, Inc. v. Ameriprise Fin., Inc.,</i>	
12	No. A-14-CA-12-SS, 2014 WL 950065 (W.D. Tex. Mar. 11, 2014).....	25
13	<i>Widespread Elec. Sales, LLC v. Upstate Breaker Wholesale Supply, Inc.,</i>	
14	No. 3:20-CV-2541-K, 2023 WL 8721435 (N.D. Tex. Dec. 17, 2023).....	8
15	<b>Statutes</b>	
16	17 U.S.C. § 101.....	<i>passim</i>
17	17 U.S.C. § 106.....	4, 7
18	17 U.S.C. § 1202.....	<i>passim</i>
19	<b>Rules</b>	
20	Fed. R. Civ P. 7.....	14
21	Fed. R. Civ P. 12.....	13, 14, 15, 20
22	<b>Treatises</b>	
23	1 Witkin, Summary of Cal. Law, Contracts § 801 (2023).....	19
24	Restatement (Second) of Contracts § 2 (1981).....	21
25	13 Williston on Contracts § 38:15.....	19
26	<b>Other Authorities</b>	
27	S. Rep. No. 105-190 (1998).....	6, 7, 9
28		



## I. INTRODUCTION

This case involves the fundamental right of software creators who license their code pursuant to open-source licenses to have their rights protected under the law. The OpenAI defendants (“OpenAI”), which knew the code was subject to these open-source licenses, copied software creators’ code to create an artificial intelligence product, Codex, that has the power to reproduce the very code on which it was trained. OpenAI could have followed the law: it could have licensed this code (as it has sought to do with others); it could have also ensured that any copies, modifications, and derivatives of this code adhered to the terms of the open-source licenses under which the original code was created. Instead, OpenAI acted with impunity.

OpenAI seeks to dismiss claims in Plaintiffs’ Second Amended Complaint (“SAC”) for violations of the Digital Millennium Copyright Act (“DMCA”) and breach of contract arising from violation of the terms of the open-source licenses (“OSLs”) that were attached to code and other materials (aka “Licensed Materials”) belonging to Plaintiffs. OpenAI’s motion under DMCA should be rejected for the same reasons the Court should reject GitHub and Microsoft’s motion to dismiss Plaintiffs’ DMCA claims. OpenAI’s attempt to dismiss the Plaintiffs’ breach of contract claim fares no better. Plaintiffs allege that the Licenses they published code subject to are valid as enforceable contracts, despite OpenAI’s arguments to the contrary.

## II. PROCEDURAL HISTORY

### A. Plaintiffs’ Initial Complaint (ECF No. 1)

On November 21, 2022, Plaintiffs Doe 1, Doe 2, Doe 3, and Doe 4 filed their initial complaint in this matter alleging, that OpenAI (and other defendants) violated DMCA §§ 1202(b)(1) and (b)(3) by removing or altering Copyright Management Information (“CMI”) from Plaintiffs’ Licensed Materials and distributing copies of those Materials that did not include CMI. ECF No. 1 ¶¶ 138–67. Plaintiffs also brought a claim for Breach of Contract—Open-Source License Violations. *Id.* ¶¶ 168–83.

**B. The Court’s Order on OpenAI’s Motion to Dismiss the Initial Complaint (ECF No. 95)**

On January 26, 2023, OpenAI filed its motion to dismiss. ECF No. 53. OpenAI argued that Plaintiffs failed to state a claim for breach of contract (ECF No. 53 at 13–14) and failed to state a claim for violation of the DMCA. ECF No. 53 at 9–13.

On May 11, 2023, the Court issued its Order. The Court upheld the sufficiency of Plaintiffs’ allegations for breach of contract and their claims under DMCA §§ 1202(b)(1) and 1202(b)(3). ECF No.95 (cited herein as *Doe 1 v. GitHub, Inc.*, 672 F. Supp. 3d 837, 859 (N.D. Cal. 2023) or “First MTD Order”). The Court also upheld the sufficiency of Plaintiffs’ breach of open-source license pleadings. *Id.* at 860.

**C. Plaintiffs’ First Amended Complaint (ECF No. 97)**

On June 8, 2023, Plaintiffs filed their First Amended Complaint. ECF No. 97. Plaintiffs maintained their claims for violations of DMCA §§ 1202(b)(1) and (b)(3), *id.* at ¶¶ 183–213, and Breach of the OSLs. *Id.* at ¶¶ 214–29.

Plaintiffs’ FAC also alleged new facts establishing Plaintiffs suffered particularized harm by specifying how Codex and Copilot output code that matches their licensed code that appears in GitHub public repositories, yet without providing the required attribution and notice and in violation of other terms of their licenses. Specifically, Plaintiffs provided examples of Copilot emitting Plaintiffs’ code verbatim as output with only cosmetic changes. *E.g., id.* at ¶¶ 100–28. The sample outputs are sufficiently distinctive to prove Copilot could have only taken the underlying code for these outputs from Plaintiffs’ GitHub repositories. *E.g., id.* at ¶¶ 103, 110, 119. As the FAC alleged, these examples illustrate the typical behavior of Copilot and raise a plausible inference that Plaintiffs’ code has also been output in the past. *Id.* at ¶ 97.

OpenAI filed its motion to dismiss the FAC on June 29, 2023. ECF No. 110. OpenAI did not attempt to dismiss Plaintiffs’ breach-of-contract claims based on the OSLs. As to the DMCA claim, OpenAI argued that Plaintiffs had failed to allege CMI removal from identical copies. ECF No. 110 at 12–13.

**D. The Court's Order on OpenAI's Motion to Dismiss First Amended Complaint (ECF No. 189)**

On January 3, 2024, the Court issued its order on the motion to dismiss the FAC. ECF No. 189 (hereinafter "Second MTD Order"). The Court dismissed Plaintiffs' claims under DMCA §§ 1202(b)(1) and (b)(3), holding that the pleadings were insufficient because output from Copilot is often a modification of Plaintiffs' licensed works, as opposed to an "identical copy." *Id.* at 15. Though the Court recognized that DMCA § 1202(b)(1) only requires intentionally removing or altering any copyright recognized information without the authority of the copyright owner, and is distinct from § 1202(b)(3), which makes it illegal to "distribute, [or] import for distribution . . . copies of works," the court applied an "identity" requirement to both DMCA §§ 1202(b)(1) and (b)(3). *Id.* at 15. The Court cited *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at \*6 (C.D. Cal. Aug. 14, 2020), a case that involved jewelry designs where the CMI was engraved in the copyrighted work. ECF No. 189 at 15. The Court still granted Plaintiffs leave to amend their DMCA §§ 1202(b)(1) and (b)(3) claims.

**E. Plaintiffs' SAC (ECF No. 200) and Its New Allegations**

On January 25, 2024, Plaintiffs filed their Second Amended Complaint. ECF No. 200. Consistent with the Second MTD Order, the SAC included new allegations regarding the increased likelihood that Plaintiffs' or class members' licensed code would be emitted verbatim as the capacity and scope of the Copilot and Codex model grows. *Id.* at ¶¶ 104–07.

**III. ARGUMENT**

**A. Plaintiffs' Second Amended Complaint Plausibly Alleges DMCA Violations under DMCA §§ 1202(b)(1) and 1202(b)(3)**

The DMCA restricts "the removal or alteration of copyright management information ("CMI")—information such as the title, the author, the copyright owner, the terms and conditions for use of the work, and other identifying information set forth in a copyright notice or conveyed in connection with the work." *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 671 (9th Cir. 2018). Sections 1202(b)(1) and (b)(3) of the DMCA, under which Plaintiffs plead, provide that



one cannot, without authority of the copyright owner, “(1) “intentionally remove or alter any copyright management information,” or (3) “distribute or import for distribution . . . copies of works . . . knowing that copyright management information has been removed or altered” while “knowing, or . . . having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement of *any right under this title*.” 17 U.S.C. § 1202(b) (emphasis added). Rights under Title 17, which encompasses the Copyright Act and DMCA, are not limited to the exclusive right of copyright holders to make copies of their works under § 106(1), but also include the exclusive right to make derivative works under § 106(2), and to distribute copies under § 106(3). *See* 17 U.S.C. § 106. Further, the Copyright Act’s definition of “copies” does not require “identity.” 17 U.S.C. § 101.

“The pleading burden under the DMCA is low.” *See Mollman v. Zoetop Bus. Co.*, No. CV 22-4128 (PA) (GJSx), 2022 WL 17207103, at \*3 (C.D. Cal. Sept. 16, 2022). Moreover, while § 1202(b) claims require a plaintiff to plead scienter, such knowledge may be alleged generally and can be inferred through circumstantial evidence of intent. *See, e.g., Logan v. Meta Platforms, Inc.*, 636 F. Supp. 3d 1052, 1063–64, 2022 WL 14813836, at \*9 (N.D. Cal. Oct. 25, 2022) (“‘[I]ntent, knowledge, and other conditions of a person’s mind may be alleged generally’” for purposes of pleading DMCA § 1202(b)’s mental state requirements”) (citations omitted).

A violation of § 1202(b)(1) is distinct from a violation of § 1202(b)(3) because it does not require the “distribution” or “import for distribution” of “copies of works” for liability. “To plead a violation of Section 1202(b)(1), a plaintiff must allege intentional removal of CMI.” *Harrington v. Pinterest, Inc.*, No. 5:20-CV-05290-EJD, 2022 WL 4348460, at \*4 (N.D. Cal. Sept. 19, 2022). The DMCA makes the mere removal of CMI from digital copies illegal before distribution of copies. To the extent an identity requirement can be read into § 1202(b)(1), it should focus on how the copy was made and CMI removed—i.e., whether it was digital copying and removal of CMI. Plaintiffs have alleged that OpenAI made direct identical digital copies of Plaintiffs’ works and thereafter automated the removal of its CMI. SAC ¶¶, 91, 94–97, 143–44.

In its First MTD Order (ECF No. 95) this Court held that Plaintiffs had sufficiently alleged violations of DMCA §§ 1202(b)(1) and 1202(b)(3) because Plaintiffs had plausibly alleged

that their “licensed code contain[ed] CMI,” and that “Defendants removed or altered that CMI from licensed code, distributed CMI knowing CMI had been removed or altered, and distributed copies of the code knowing that CMI had been removed or altered, all while knowing and possessing reasonable grounds to know that doing so would induce infringement.” *Doe 1*, 672 F. Supp. 3d 837, 857.

In its Second MTD Order, however, the Court dismissed Plaintiffs’ DMCA claims solely on the grounds that “no DMCA violation exists where the works are not identical,” and Plaintiffs had failed to satisfy this requirement “[b]ecause the Complaint characterized Copilot output as ‘a “modified format,” ‘variation[ ],’ or the ‘functional[ ] equivalent’ of the licensed code.” ECF No. 189 at 15 (quoting FAC ¶¶ 103, 110, 120).

In sum, this Court has already held that Plaintiffs have otherwise sufficiently pled DMCA violations under §§ 1202(b)(1) and 1202(b)(3). The sole issue remaining before the Court is whether Plaintiffs’ SAC sufficiently alleges that Copilot produces “identical” copies of Plaintiffs’ licensed code. For the following reasons, and for the reasons the Court set forth in its First MTD Order, Plaintiffs’ SAC satisfies this requirement, and thus Plaintiffs’ claims under §§ 1202(b)(1) and 1202(b)(3) for injunctive relief should not be dismissed.

# **1. Plaintiffs’ DMCA claims are plausibly alleged under any standard**

Plaintiffs do not seek to relitigate previous arguments concerning the DMCA’s alleged “identity” requirement and its application to the facts in this case. Plaintiffs, however, ask this Court to address new issues arising in relation to this standard and its application to software code in particular—issues that were either never previously raised or addressed in prior briefing or the Court’s Second MTD Order.<sup>1</sup>

Relatedly, Plaintiffs ask this Court to reconsider the caselaw on which it relied in reaching its “identity” holding in the Second MTD Order, particularly in light of the recent district

---

<sup>1</sup> ECF No. 189 (“A court may consider a motion “to the extent it presents issues not previously resolved.” (quoting *Jones v. Life Ins. Co. of N. Am.*, No. 08-CV-03971-RMW, 2015 WL 8753996, at \*3 (N.D. Cal. Dec. 15, 2015)). While Plaintiffs acknowledge that there is a pending motion for reconsideration of the Court’s Order on the FAC MTD, the arguments asserted herein also differ from the arguments raised there.



1 court decision in *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411 (S.D. Tex.  
2 2023) (“*ADR Int’l*”), where this alleged “identity” element was squarely rejected. *See id.* at  
3 427–29 (finding no “identity” requirement for DMCA claims after examining the statutory  
4 language, legislative history, and caselaw).

5 **a. Identity is not an element of a DMCA § 1202 claim**

6 The plain language of DMCA § 1202 makes it a violation to remove or alter CMI. It does  
7 not require that the output work be original or identical to obtain relief. *See* 17 U.S.C.  
8 § 1202(b)(1). Similarly, the definition of “copy” under the DMCA includes no requirement of  
9 identity. *See* 17 U.S.C. § 1202(c). Indeed, the DMCA contains only one mention of the word  
10 “identical”—as an exemption under Section 1201 for certain types of institutions. *See* 17 U.S.C.  
11 § 1201(d)(2) (“The exemption made available under paragraph (1) shall only apply with respect  
12 to a work when an *identical* copy of that work is not reasonably available in another form.”)  
13 (emphasis added). One can presume therefore that Congress could have required an “identical  
14 copy” element into other sections of the DMCA if it so intended.

15 To be considered a copy under the DMCA, the “allegedly infringing work must be fixed in  
16 some tangible form from which the work can be perceived, reproduced, or otherwise  
17 communicated, either directly or with the aid of a machine or device.” *Columbia Pictures Indus.,*  
18 *Inc. v. Galindo*, No. 20-cv-3129-MEMF (GJSx), 2022 WL 17094713, at \*8 (C.D. Cal. Nov. 18,  
19 2022) (citing 17 U.S.C. § 101). The statute defines CMI as “information conveyed in connection  
20 with copies . . . of a work[.]” 17 U.S.C. § 1202(c). By a plain reading of the statute, there is no  
21 need for a copy to be identical—there only needs to be copying, which Plaintiffs have amply  
22 alleged. *See, e.g.*, SAC ¶¶ 112–40.

23 The DMCA was enacted to protect copyrights in the digital era where digital copies can  
24 be made and disseminated in an instant. *See, e.g., Textile Secrets Int’l, Inc. v. Ya-Ya Brand Inc.*, 524  
25 F. Supp. 2d 1184, 1199 (C.D. Cal. 2007) (“With this evolution in technology, the law must adapt  
26 in order to make digital networks safe places to disseminate and exploit copyrighted works.”); *see*  
27 *also* S. Rep. No. 105-190, at 2 (1998) (“[T]he law must adapt in order to make digital networks  
28 safe places to disseminate and exploit copyrighted materials”) (footnote call numbers omitted).

Here, Plaintiffs allege that OpenAI utilized identical digital copies of their copyrighted works to train Codex (SAC, ¶¶ 91, 95–97) and removed the CMI from it knowing that Copilot (which runs on top of Codex and which OpenAI codeveloped with GitHub, *id.* at ¶ 59) reproduces Plaintiffs’ copyrighted code identically, without attribution, notice, or any of the other requirements that may adhere from the applicable license. *Id.* at ¶¶ 102–03, 145–49. OpenAI knowingly enables and induces Copilot users to infringe on Plaintiffs’ exclusive right under § 106(2) of the Copyright Act to prepare derivative works. 17 U.S.C. § 106(2); *see* SAC, ¶¶ 90–105. Liability is established under DMCA § 1202 (b)(1) when CMI is removed from digital copies knowing doing so “will induce, enable, facilitate, or conceal an infringement of *any right* under this title.” 17 U.S.C. § 1202(b).

*None* of the cases that Defendants previously relied on—and this Court cited in its Second MTD Order adopting an “identity” standard for DMCA violations—explain the origins of this so-called standard. For good reason—there is no support from statutory language, legislative history, or potentially analogous interpretations of “copying” under the Copyright Act. In fact, careful review of many of these cases also cite other cases that never discussed the issue of “identity” (despite being cited for that proposition).

For example, this Court (and Defendants) cited *Kirk Kara Corp. v. W. Stone & Metal Corp.*, No. CV 20-1931-DMG (EX), 2020 WL 5991503, at \*6 (C.D. Cal. Aug. 14, 2020) as central support for the DMCA’s alleged “identity” requirement. *Kirk Kara*, however, never examined the standard, and instead relied on three cases, each of which provide no support for it. As the court in *ADR Int’l* noted, the *Kirk Kara* court cited *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) and *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. CIV 13-00496 SOM, 2015 WL 263556, at \*2 (D. Haw. Jan. 21, 2015), “but neither [of these cases] mentioned nor employed an identical copies requirement.” *ADR Int’l*, 667 F. Supp. 3d. at 427. The *Kirk Kara* court similarly cited *Fischer v. Forrest*, 286 F. Supp. 3d 590 (S.D.N.Y. 2018), yet the term “identical” appears nowhere in the decision. To the contrary, in the *Fischer* court’s brief discussion of whether sufficient copying had occurred, the court did not apply an identity standard at all, stating that where DMCA claims had been adequate, “the underlying work has



1 been *substantially* or entirely reproduced.” *Id.* at 609 (emphasis added); *see also ADR Int’l*, 667 F.  
2 Supp. 3d at 427 (rejecting *Kirk Kara*’s reliance on *Fischer*).

3 Notably, in *ADR Int’l*, the district court expressly rejected OpenAI’s claim that the  
4 DMCA included an “identity” requirement in part by dissecting the very cases on which  
5 OpenAI here—and this Court in its Second MTD Order—relied on when imposing an  
6 “identity” requirement on DMCA violations. *ADR Int’l*, 667 F. Supp. 3d at 427–29  
7 (observing that “Defendants’ cases do not require a plaintiff to plead allegedly infringing works  
8 are identical copies” and recognizing “the one case that requires identical copies [i.e., *Kirk Kara*]  
9 is not well reasoned.”).

10 The *ADR Int’l* court also held that neither the statute’s plain language nor its legislative  
11 history—which the district court examined (unlike the cases that OpenAI relies on)—supported  
12 an “identity” requirement. Instead, well-established standards of statutory construction  
13 compelled the opposite conclusion, namely, that “copying” under the DMCA should align with  
14 “copying” under the Copyright Act, given that the DMCA is contained in Chapter 12 of the  
15 Copyright Act and, therefore, subject to Section 101 of the Copyright Act’s definition of copying.  
16 Section 101 defines “copies” as “material objects, other than phonorecords, in which a work is  
17 fixed by any method now known or later developed, and from which the work can be perceived,  
18 reproduced, or otherwise communicated, either directly or with the aid of a machine or device.”  
19 17 U.S.C. § 101. This definition “lacks any requirement for an identical copy.” *ADR Int’l*, 667 F.  
20 Supp. 3d at 427; *cf. Murphy v. Millennium Radio Grp. LLC*, 650 F.3d 295, 305 (3d Cir. 2011)  
21 (“Defendants are essentially asking us to rewrite § 1202 to insert a term—that is, ‘automated  
22 copyright protection or management system’—which appears nowhere in the text of the DMCA  
23 and which lacks a clear definition. We would need compelling justification indeed to adopt such a  
24 statutorily-unmoored interpretation.”).

25 In accord with this correct analysis, courts have routinely declined to require identity  
26 when evaluating DMCA claims under a standard drawn directly from Copyright Act infringement  
27 claims. *See, e.g., Friedman v. Live Nation Merch., Inc.*, 833 F.3d 1180, 1188 (9th Cir. 2016);  
28 *Widespread Elec. Sales, LLC v. Upstate Breaker Wholesale Supply, Inc.*, No. 3:20-CV-2541-K, 2023

WL 8721435, at \*1 (N.D. Tex. Dec. 17, 2023); *Software Pricing Partners, LLC v. Geisman*, No. 3:19-cv-00195-RJC-DCK, 2022 WL 3971292, at \*5 (W.D.N.C. Aug. 31, 2022) (“Geisman, as a former employee of SPP, reasonably knew that such information was copyrighted work and knew he was altering it by changing it enough to look like his own work while maintaining a substantial similarity to the original work. Accordingly, Geisman is liable for violating the DMCA.”). Under the DMCA, all these actions would fail if courts adhered to the “identity” standard OpenAI proposes—and which this Court adopted in the Second MTD Order. OpenAI has not cited a single case holding that the standard for “copying” under the DMCA is higher than the Copyright Act or that the statutes should be read disharmoniously.

In addition to being incorrect, adopting OpenAI’s rigid “identity” standard to software code also has uniquely troubling implications given that mere cosmetic differences between the original and copied code are not probative of whether the copied code is “identical.” *See* 17 U.S.C. § 1201(d)(2) (containing the only mention of the word “identical” in 17 U.S.C. § 1201(d)(2) regarding an exemption for public institutions); *see also* S. Rep. No. 105-190, at 31 (1998) (explaining the meaning of public institution exemption and requirement that copies be “identical” for exemption to apply). The decision in *Oracle Int’l Corp. v. Rimini St., Inc.*, No. 2:14-cv-01699-MMD-DJA, 2023 WL 4706127 (D. Nev. July 24, 2023) is especially instructive. There, the court rejected the party’s argument “that a work that removes copyright management information must be an *exact copy of the original work*.” *Id.* at \*82 (emphasis added). The court stated that “this construction of the DMCA” demanding an exact copy—the same construction OpenAI demands in this case—“would weaken the statute’s intended protections for copyright holders” and that “when a defendant modifies source code ‘substantially similar’” to the underlying “copyrighted source code, including by replacing the author’s name with its own, the defendant is liable under the DMCA.” *Id.* (internal quotations omitted); *Rimini St., Inc.*, 2023 WL 4706127, at \*14 (examining whether copies of code were “substantially similar”). In a case involving software code in this very district, the court found that an allegation of CMI removal from code that is derivative of a plaintiff’s copyrighted source code is sufficient at the pleading stage to sustain a DMCA § 1202 claim. *See Splunk Inc. v. Cribl, Inc.*, 662 F. Supp. 3d 1029, 1053



(N.D. Cal. 2023). Identity is simply not an element of a Section 1202 claim.

**b. Plaintiffs allege that Copilot reproduces identical copies of Plaintiffs' code**

Even if “identity” were actually an element of a DMCA violation, Plaintiffs’ allegations under §§ 1202(b)(1) and (b)(3) still satisfy this standard. Plaintiffs have plausibly alleged that any differences between Copilot-generated code and Plaintiffs’ original code are “immaterial” and “semantically insignificant”—in other words: identical. *See, e.g.*, SAC ¶ 108.

Plaintiffs identify several instances where Copilot *has* reproduced verbatim, (i.e., identical) copies of Plaintiffs’ code. *See* SAC ¶¶ 135, 137. Plaintiffs note that “[w]hen Copilot is prompted with the first section of Doe 5’s code . . . [t]he first suggestion from Copilot offers . . . a verbatim copy of Doe 5’s original code.” *Id.* at ¶ 135. Plaintiffs similarly allege that Copilot reproduces “a verbatim copy of Doe 5’s code (except for small cosmetic variations in line breaks)” when prompted by different lines of Doe 5’s code. *Id.* at ¶ 137. Line breaks do not comprise the copyrighted work as they are not “semantically meaningful,” *id.* at ¶ 72, and thus any difference in line breaks between Copilot output and Plaintiffs’ code are immaterial to the analysis.

Elsewhere, the SAC plausibly alleges that Copilot reproduces code sufficient to satisfy an “identity” requirement, because minor deviations between the original code and Copilot’s copy of that code are “immaterial” or “semantically insignificant variations of the original Licensed Materials.” *Id.* at ¶ 108. With respect to Doe 2, Plaintiffs allege that slight differences in three key words still render the Copilot code “semantically equivalent” and that the code is otherwise “a verbatim copy” of the original code. *Id.* at ¶ 115. Plaintiffs further allege that “the particular arrangement and sequencing seen in [Doe 2’s] code is distinctive expression found only in one location on GitHub.” *Id.* Similarly, Plaintiffs allege that Copilot clearly reproduces Doe 1’s code given that its omitted output only contains “cosmetic” differences that do not affect the way the code functions, and that the output even contains Doe 1’s actual “comments in the code”—in other words, the coder’s own expressive commentary about his own code, which could only come from Doe 1, further evidencing copying. *Id.* at ¶ 121.

Plaintiffs further allege that Copilot renders a “verbatim copy” of Doe 5’s original code except with an aspect of the code that is immaterial to the question of whether it has been copied. *Id.* ¶ 128. The renaming of variables is akin to renaming James Bond as “Jim Bond”: republishing the popular 007 movie series would still constitute an act of infringement irrespective of the new names. Indeed, the SAC is replete with allegations making clear that the scant “cosmetic” differences in Copilot-generated code have no bearing on whether the Copilot-generated code is “functionally equivalent” or identical to Plaintiffs’ code. *See, e.g., id.* at ¶¶ 122, 131, 137. Plaintiffs also allege that these small cosmetic differences are “a feature, not a bug” that enables “Copilot to conceal the copying of Licensed Materials.” *Id.* at ¶ 155.

OpenAI’s contention that minor differences in line breaks, capitalization, or variable names defeat a finding of identity is a factual argument that can be tested through expert testimony showing that these modifications are, in fact, material such that the code is, in fact, *not* identical. *See, e.g., Rimini St., Inc.*, 2023 WL 4706127, at \*14 (“Ms. Frederiksen-Cross performed analytic dissection as part of her comparison of RSI810ST.SQR against TAX810DC.SQR. She determined that the RSI810ST.SQR file contains substantial portions of Oracle’s protected expression and was substantially similar to Oracle’s TAX810DC.SQR file. Examples of Oracle’s protected expression within the RSI810ST.SQR file include, “blocks of identical lines that are performing exactly the same function using identical code in the same order.”) (internal citations omitted). But this case is about *code*. Minor differences may have no implication as to whether the code functions identically or not.

In addition, Plaintiffs’ SAC contains new allegations regarding Copilot’s duplicate-detection feature, which GitHub created “in response to public criticism of Copilot’s mishandling of Licensed Materials.” *Id.* at ¶ 145. As Plaintiffs allege, the feature “either allow[s] or block[s] code completion suggestions that match publicly available code.” *Id.* In other words, the feature was created because GitHub, OpenAI, and the other defendants *knew* that Copilot was emitting verbatim copies of licensed code on which it had trained. *See id.* at ¶¶ 144–48. A tool that attempts to block Copilot from producing output that is identical to code hosted in public GitHub repositories *by definition* establishes Copilot’s ability to reproduce verbatim copies of code.



Moreover, as Plaintiffs allege, “even assuming the filter works as advertised, because it only checks for verbatim excerpts, it does nothing to impede the Outputs from Copilot that are modifications of Licensed Materials.” *Id.* at ¶ 145. In other words, code that includes minor differences in line breaks, capitalization, or even order of enums—all semantically “immaterial differences”—are not necessarily flagged by the feature. *Id.*

At the pleading stage, Plaintiffs’ allegations regarding the duplicate-detection feature clearly bolster the plausibility of Plaintiffs’ claims that Copilot produces identical code and is substantially likely to continue producing Plaintiffs’ licensed code.

## 2. The DMCA does not require an entire work to be copied

OpenAI also claims, without citation to a single authority, that “the reproduction of short passages that may be part of [a] larger work, rather than the reproduction of an entire work, is insufficient to violate [DMCA] Section 1202.” OpenAI Br. at 5. But the DMCA has no such requirement.

Like OpenAI’s “identity” argument, its claim that DMCA liability attaches only where the offending party has copied an “entire work” contravenes the most basic tenets of copyright law. By OpenAI’s logic, a party could copy and distribute a fragment of a copyrighted work—say, a chapter of a book, a stanza of a poem, or a scene from a movie—and face no repercussions for infringement. Courts, however, have routinely acknowledged that “[e]ven a small amount of the original, if it is qualitatively significant, may be sufficient to be an infringement” even where “the full original could not be recreated from the excerpt.” *Horgan v. Macmillan, Inc.*, 789 F.2d 157, 162 (2d Cir. 1986); *see also Burroughs v. Metro-Goldwyn-Mayer, Inc.*, 683 F.2d 610, 624 n.14 (2d Cir. 1982).

OpenAI’s unsupported theory would render the DMCA toothless by allowing anyone to copy another’s work so long as they only copied a part of it. But “[n]o plagiarist can excuse the wrong by showing how much of his work he did not pirate.” *Shaw v. Lindheim*, 919 F.2d 1353, 1362 (9th Cir. 1990), *overruled by Skidmore v. Led Zeppelin*, 952 F.3d 1051 (9th Cir. 2020).

**B. Plaintiffs Adequately Allege Breach of Contract by OpenAI Through Its Mass Violation of Open-Source Licenses**

In its initial Motion to Dismiss, OpenAI urged the Court to dismiss Plaintiffs’ breach of contract claim (Count II), contending that Plaintiffs had failed to plead the existence of a contract and failed to identify the actual contractual provisions OpenAI had breached. ECF No. 53 at 13–15. The Court rejected these arguments wholesale, holding Plaintiffs had sufficiently alleged “that Codex and Copilot reproduce licensed code as output without attribution, copyright notice, or license terms, thereby violating the relevant provisions of each license,” and that while Plaintiffs had not identified “the specific subsections of each suggested license that correspond to each of these requirements,” they had “sufficiently identified ‘the contractual obligations allegedly breached,’ as required to allege breach of contract. *Doe 1*, 672 F. Supp. 3d at 860.

Plaintiffs then filed their FAC, leaving their breach-of-contract allegations unchanged. OpenAI did not move to dismiss this count. ECF No. 110. In the SAC, the same nucleus of facts underlying Plaintiffs’ breach of contract claim remain unchanged. After filing the SAC—which contains the very same breach of contract claim upheld by this Court in its First MTD Order—OpenAI sought another bite at the apple, advancing two central claims for the first time.

First—that Plaintiffs have failed to allege a breach based on the operation of OpenAI’s Codex and thus have failed to state a claim against OpenAI based on Codex given that “GitHub” is allegedly a “separate product” from Copilot. OpenAI Br. at 7–8.

Second—that Plaintiffs have failed to articulate a theory of breach of contract because OpenAI never adhered to the licensing “conditions” to begin with, meaning its use of the Plaintiffs’ Licensed Materials was never in fact governed by the attached licenses. *Id.* at 10.

For the reasons below, OpenAI’s arguments fail.

**1. OpenAI’s failure to raise these arguments before forecloses OpenAI’s Rule 12 motion under Rule 12(g)(2)**

“Rule 12(g)(2) provides that a defendant who fails to assert a failure-to-state-a-claim defense in a pre-answer Rule 12 motion cannot assert that defense in a later pre-answer motion under Rule 12(b)(6)” though “the defense may be asserted in other ways.” *In re Apple iPhone*



1 *Antitrust Litig.*, 846 F.3d 313, 318 (9th Cir. 2017) (citing Fed. R. Civ. P. 12(g)(2) & (h)(2)); *see also*  
2 Fed. R. Civ. P. 12(g)(2) (“Except as provided in Rule 12(h)(2) or (3), a party that makes a motion  
3 under this rule must not make another motion under this rule raising a defense or objection that  
4 was available to the party but omitted from its earlier motion.”). “The sole Rule 12 exception is  
5 that a party may subsequently raise the foreclosed issue ‘in a post-answer motion under  
6 Rule 12(c)’; otherwise, the party may validly raise the issue ‘in a pleading under Rule 7 . . . or at  
7 trial.” *In re Packaged Seafood Prods. Antitrust Litig.*, 277 F. Supp. 3d 1167, 1174 (S.D. Cal. 2017)  
8 (quoting *Apple iPhone*, 846 F.3d at 318). Moreover, when assessing a 12(g)(2) motion, district  
9 courts consider the prejudice to plaintiffs in allowing a defendant to seek dismissal of claims on  
10 grounds that could have been brought through an earlier motion. *See, e.g., Harrell v. City of Gilroy*,  
11 No. 17-CV-05204-LHK, 2019 WL 452039, at \*8 (N.D. Cal. Feb. 5, 2019) (noting that courts have  
12 discretion to consider successive Rule 12(b)(6) motions if the motion does not prejudice the  
13 plaintiff and expedites resolution of the case.”).

14 The facts in this case are precisely why Rule 12(g)(2) exists: “to avoid repetitive motion  
15 practice, delay, [or] ambush tactics.” *Packaged Seafood Prods.*, 277 F. Supp. 3d at 1174 (quotations  
16 omitted and alteration in the original). This is OpenAI’s *third* motion to dismiss; yet, its *first* time  
17 arguing for a Rule 12 dismissal of Plaintiffs’ breach of contract claims—which have not changed  
18 one iota in any of the complaints. *Compare* ECF No. 201 *with* ECF No. 1. OpenAI has known of  
19 these allegations for nearly 16 months, when Plaintiffs filed their initial Complaint. ECF No. 1.  
20 Nothing prevented OpenAI from making the exact same breach of contract arguments it pursues  
21 today, and OpenAI has offered no explanation as to why it failed to raise these argument earlier.  
22 Nor has OpenAI explained why it failed to raise these arguments in its second motion to dismiss,  
23 instead sitting silent until its third round of Rule 12(b)(6) briefing. *See Mario V. v. Armenta*, No.  
24 18-cv-00041-BLF, 2019 WL 81337140, at \*2 (N.D. Cal. Apr. 17, 2019) (denying subsequent Rule  
25 12(b)(6) motion under Rule 12(g)(2) where “Plaintiffs filed their original complaint more than a  
26 year ago,” “Defendants’ first motions to dismiss were litigated more than eight months ago” and  
27 “[t]he issues now raised by Defendants could have, and under the Federal Rules should have,  
28 been litigated then.”).



OpenAI should not be rewarded for its delay in advancing breach of contract arguments that could have been made on two prior occasions and, notably, before discovery had commenced. At this point in the case, however, Plaintiffs have issued numerous discovery requests pertaining to their breach of contract claims, met and conferred on these issues in person, exchanged six letters in relation to discovery, and also filed one joint letter brief with the Court. Indulging OpenAI's belated attempt to dismiss Plaintiffs' breach of contract claim at this stage in the litigation would thus be manifestly unjust to Plaintiffs. *See Packaged Seafood Prods.*, 277 F. Supp. 3d at 1174 (“[T]o refuse to enforce Rule 12(g)(2)’s clear command on such a foundational argument as the one Defendants here urge . . . would set a dangerous precedent regarding the ability to continually hamstring a plaintiff with wave after wave of motions to dismiss.”).

Notwithstanding that OpenAI has waived its right challenge Plaintiffs' breach of contract claim, its arguments fail on the merits, as well.

**2. Plaintiffs adequately allege breach of contract based on the operation of OpenAI's Codex product and its integration into GitHub and OpenAI's joint venture, the Copilot product**

OpenAI incorrectly contends that Plaintiffs' breach-of-contract claim should be dismissed as to OpenAI specifically, on the grounds that Plaintiffs' claim only concerns the operation of Copilot and not Codex, and OpenAI bases this assertion that Plaintiffs “acknowledge that OpenAI Codex and GitHub Copilot are different products by different parties.” OpenAI Br. at 7. OpenAI's characterization of Codex and Copilot as “different products” made by “different parties” ignores countless allegations in the SAC plausibly alleging a deeply integrated relationship between these products. It also ignores the allegation of a joint venture relationship between OpenAI and GitHub/Microsoft pursuant to which OpenAI has joint and several liability for the claims alleged in the Complaint. SAC ¶ 59; *see also id.* at ¶ 27 (alleging that OpenAI “co-created Copilot and offers it jointly with GitHub”); ¶ 45 (“Each acted as the principal, agent, or joint venture of, or for other Defendants with respect to the acts, violations, and common course of conduct herein alleged”). Both OpenAI's copying and use of Plaintiffs' licensed code through Codex, and Copilot's Codex-based reproduction of Plaintiffs' licensed code gives rise to breach-of-contract claims. *See* SAC ¶ 68 (stating that Codex “has not been trained to provide Attribution.

Nor does it include a Copyright Notice nor any License Terms attached to the Output. This is by design”); *id.* at 11 (describing breach of contract claims in “Contract-Related Conduct” section in terms of “[d]efendants violat[ing] the Licenses governing the use of the Licensed Materials by using them to train Copilot and for republishing those materials without appending the required Attribution, Copyright Notice, or License Terms.”)

Contrary to OpenAI’s claims, Plaintiffs allege that “Codex and Copilot are related” precisely because Codex “powers GitHub Copilot, which [OpenAI] built and launched in partnership with GitHub.” SAC ¶ 59. Specifically, “GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time.” *Id.* at ¶ 47 (quoting GitHub website). Plaintiffs have additionally alleged that both Codex and Copilot fail to identify “any Licensing Terms attached to the Output” as neither was “programmed to treat attribution, copyright notices, and license terms as legally essential.” *Id.* at ¶¶ 68, 92. Moreover, Plaintiffs allege that “Defendants have kept secret the details of Codex’s modifications and its integration into or interaction with Copilot,” *id.* at ¶ 59, which this Court acknowledged in its First MTD Order, *Doe 1*, 672 F. Supp. 3d at 851 n.9 (Defendants are “the only parties with knowledge of how Copilot and Codex were designed and operate.”). Defendants’ lack of transparency underscores the need for discovery on this very point.

In addition to Plaintiffs’ allegations regarding Codex’s integration into Copilot, Plaintiffs specifically allege that GitHub and OpenAI “launched Copilot in June 2021” and that they “co-created Copilot” as a joint venture and jointly share in its profits. SAC ¶¶ 26–27, 45, 59, 225. Parties to a joint venture are jointly and severally liable for the actions of their partners. *See, e.g., Myrick v. Mastagni*, 185 Cal. App. 4th 1082, 1091 (2010) (“The incidents of a joint venture are in all important respects the same as those of a partnership. One such incident of partnership is that all partners are jointly and severally liable for partnership obligations, irrespective of their individual partnership interests.” (quoting 9 Witkin, Summary of Cal. Law (10th ed. 2005) Partnership, § 9, p. 584.)). And “where there is a joint venture, all members of a joint venture are liable for contracts executed by another member in furtherance of the joint venture.” *Forest v. Equitable Life Assurance Soc’y of U.S.*, No. C99-5173 SI, 2001 WL 1338809, at \*4 (N.D. Cal. June



12, 2001) (collecting cases).

**3. OpenAI’s failure to abide by open-source licenses adequately supports a breach-of-contract claim**

Although this Court already upheld Plaintiffs’ breach-of-contract claim in its First MTD Order, OpenAI now argues that this claim should be dismissed because the attribution and notice requirements imposed by Plaintiffs’ Suggested Licenses are conditions of the copyright license rather than covenants. OpenAI Br. at 10. OpenAI claims that “failure to meet a condition does not give rise to a claim for breach of contract.” *Id.* at 18. Thus, if “Plaintiffs are right that there has been a failure to fulfill the attribution and notice terms,” OpenAI claims, the consequence of that failure is that no license was granted, and Plaintiffs may only seek redress through a copyright infringement action brought under the Copyright Act. *Id.* at 10, 26.

As discussed below, OpenAI is plainly wrong on the law. That the Suggested Licenses create an enforceable contract is not novel. The facts giving rise to this claim are straightforward—indeed, this Court has recognized that Plaintiffs have pleaded all elements of a breach of contract claim based on the Suggested Licenses. *Doe 1*, 672 F. Supp. 3d at 859–60. This Court is not alone in upholding breach of contract claims based on OSLs—several courts have upheld breach of contract claims involving OSLs, including some of the very Suggested Licenses at issue in this case. *See, e.g., Software Freedom Conservancy, Inc. v. Vizio, Inc.*, No. 8:21-cv-01943-JLS-KES, 2022 WL 1527518, at\*3 (C.D. Cal. May 13, 2022) (“*SFC*”); *Artifex Software, Inc. v. Hancom, Inc.*, No. 16-CV-06982-JSC, 2017 WL 1477373, at \*3 (N.D. Cal. Apr. 25, 2017). Instead, OpenAI’s positions would appear to upset any claim that these OSLs are valid or enforceable. As in accord with how these courts have concluded, that is wrong.

OpenAI’s position also ignores that Plaintiffs have alleged far more than breaches of attribution and notice requirements. *See, e.g., SAC ¶¶ 244, 247.* Thus, even if OpenAI is correct about its willful breach of attribution and notice terms in the OSLs at issue in this case (it is not), at most this would warrant a narrowing of Plaintiffs’ breach of contract claim—not dismissal of it.



**a. Plaintiffs sufficiently plead breach of contract because the attribution and notice requirements of OSLs are properly considered covenants and not conditions**

OpenAI argues that attribution and notification duties are “conditions” of the grant of the copyright license, breach of which can only be remedied by a copyright infringement claim rather than pursuant to a breach of contract claim. OpenAI Br. at 10–19. Make no mistake—OpenAI’s ambition here is to utterly eviscerate open-source licensing in the large, including the OSLs that required it to provide attribution and notice of the source code it used to create Codex and Copilot, as well as provide such attribution and notice when those models reproduce identical copies of Plaintiffs’ source code. *Id.*

As a baseline, a “copyright owner who grants a nonexclusive, limited license ordinarily waives the right to sue licensees for copyright infringement, and [] may sue only for breach of contract.” *Ticketmaster L.L.C. v. Prestige Ent., Inc.*, 306 F. Supp. 3d 1164, 1172 (C.D. Cal. 2018) (quoting *Sun Microsystems, Inc. v. Microsoft Corp.*, 188 F.3d 1115, 1121 (9th Cir. 1999)) (internal quotations omitted). “However, if the licensee acts outside the scope of the license, the licensor may sue for copyright infringement.” *Id.* (quoting *MDY Indus., LLC v. Blizzard Ent., Inc.*, 629 F.3d 928, 939 (9th Cir. 2010)). The Ninth Circuit has referred to contractual terms that limit a copyright license’s scope as “conditions,” the breach of which constitute copyright infringement, but it refers to all other license terms as “covenants,” the breach of which is actionable only under contract law. *MDY Indus.*, 629 F.3d at 939, *as amended on denial of reh’g* (Feb. 17, 2011), *opinion amended and superseded on denial of reh’g*, No. 09-15932, 2011 WL 538748 (9th Cir. Feb. 17, 2011). The distinction between conditions and covenants is decided “according to state contract law” and “[w]herever possible, equity construes ambiguous contract provisions as covenants rather than conditions.” *Id.*

Under some interpretations of contract law, “conditions” relate to mutually agreed limitations on the scope of the licensor’s limited use of the copyrighted material. “A condition is created by the *mutual* assent of the parties and is binding on both, while a covenant is binding on the covenantor only.” *Ticketmaster*, 306 F. Supp. 3d at 1172–73 (emphasis added). Under California law, contractual provisions can also operate as both conditions and covenants, *see, e.g.*,

1 *Call v. Alcan Pac. Co.*, 251 Cal.App.2d 442, 447 (1967), and parties can either treat the contract as  
2 terminated or enforce the covenant, *see, e.g.*, 1 Witkin, Summary of Cal. Law, Contracts § 801  
3 (2023) (“The same fact or act can be both a condition and a promise”); 13 Williston on Contracts  
4 § 38:15 (collecting cases). Importantly, describing terms in a license as “conditions” does *not*  
5 control whether a term is treated as a condition or covenant under contract law. Instead, whether  
6 a term may serve as a condition, limiting the scope of a license, depends on “the whole contract,  
7 its purpose, and the intention of the parties” regardless of whether it is referred to as a  
8 “condition.” *Ticketmaster*, 306 F. Supp. 3d at 1173 (*quoting JMR Constr. Corp. v. Env’l Assessment*  
9 *& Remediation Mgmt., Inc.*, 243 Cal. App. 4th 571, 596 (2015)); *see also Pac. Allied v. Century Steel*  
10 *Prods.*, 162 Cal. App. 2d 70, 79–80 (1958) (conditions are disfavored by the law). Generally,  
11 “[c]onditions precedent are disfavored and will not be read into a contract unless required by  
12 plain, unambiguous language.” *Netbula, LLC v. Storage Tech. Corp.*, No. C06-07391 MJJ, 2008  
13 WL 228036, at \*3 (N.D. Cal. Jan. 18, 2008) (*quoting Effects Associates, Inc. v. Cohen*, 908 F.2d 555,  
14 559 n.7 (9th Cir. 1990)).

15 OpenAI’s brief includes a multipage table that quotes extensively from some of the OSLs  
16 at issue in the SAC. OpenAI Br. at 11–17. The length of OpenAI’s table is meant to distract from  
17 its cramped and conclusory legal argument: that the mere use of the magic language “provided  
18 that” and “condition” in every case necessarily creates terms that are legally conditions, not  
19 covenants. OpenAI Br. at 11, 16. Not so.

20 In the first instance, OpenAI’s blanket assertions about the meaning of the world’s most  
21 popular OSLs—all of which are at issue in this case—would imperil the entire ecosystem of  
22 open-source licensing. OpenAI’s position that every OSL merely imposes conditions, not  
23 covenants, would mean that software coders have no redress for violation of their OSLs *unless*  
24 they have registered their code with the Copyright Office and bring a copyright-infringement  
25 action in federal court. Indeed, Plaintiffs suspect that GitHub and Microsoft have not sought to  
26 join OpenAI’s argument on this issue because it would cause a four-alarm public-relations  
27 firestorm among the millions of open-source programmers who have relied on GitHub’s  
28 representations that OSLs are meaningful, not illusory. SAC at 10, n.4 (enumerating the 11



“Suggested Licenses” that are at issue in the SAC and form part of the definition of the proposed class).

OpenAI is likely correct that as part of this litigation, the meaning of the Suggested Licenses in the context of other admissible evidence under California contract law will need to be determined. But not on the simplistic and incomplete basis OpenAI proposes. Rather, the project will require this Court to consider the text of the licenses separately, and interpret “the whole contract, its purpose, and the intention of the parties.” *Ticketmaster*, 306 F. Supp. 3d at 1173. Common evidence will likely be relevant, but ultimately, the “intention of the parties” is an issue of fact that cannot be resolved as part of a motion to dismiss, especially because all inferences must still be drawn in Plaintiffs’ favor. *Laborers Health Welf. Tr. Fund v. Kaufman & Broad of N. Cal., Inc.*, 707 F.2d 412, 418 (9th Cir. 1983) (contractual “intent . . . is a question for the trier of fact”); *McShannock*, 976 F.3d at 886–87. Plaintiffs specifically allege facts showing Plaintiffs’ intent that the Suggested Licenses create enforceable contracts to relevant license terms. OpenAI improperly seeks to foreclose this whole process, encouraging this Court to adopt a permanent interpretation of these OSLs based on selective quotations in its Rule 12 motion. OpenAI Br. at 11–17. As a procedural and substantive matter, this invitation should be declined.

**a. *Jacobsen v. Katzer* is not determinative**

The main case OpenAI cites in support of its position is *Jacobsen v. Katzer*, 535 F.3d 1373 (Fed. Cir. 2008).<sup>2</sup> In *Jacobsen*, the plaintiff, who had registered his code with the Copyright Office, offered software under an open-source license called the Artistic License. *Id.* at 1375–76. The central issue on appeal was not whether plaintiff there had adequately pled a *breach of contract claim*, but whether the district court erred in dismissing plaintiff’s *copyright infringement claim*. *Id.* at 1377. In the words of the *Jacobsen* court, “[t]he heart of the argument on appeal concerns whether the terms of the Artistic License are conditions of, or merely covenants to, the copyright license,” because “if the terms of the Artistic License allegedly violated are both covenants and conditions, they may serve to limit the scope of the license and are governed by contract law.” *Id.*

---

<sup>2</sup> Plaintiffs note that decisions of the Federal Circuit are not binding.



at 1380. If the terms were “merely covenants, by contrast, they are governed by contract law.” *Id.*

Specifically, the court analyzed whether the attribution and notice requirements in the Artistic License defendants had allegedly violated were “conditions of, or merely covenants to, the copyright license.” *Id.* at 1381. The appeals court held that the requirements were “conditions to the scope” of the license. *Id.* at 1382. But critical to the *Jacobsen* court’s holding were contractual features specific to the Artistic License—namely, that the attribution and notice requirements had to be harmonized with the “Preamble” of the Artistic License, which expressly recognized that “[t]he intent of th[e] document [wa]s to state the conditions *under which* a Package may be copied.” *Jacobsen*, 535 F.3d at 1381 (emphasis added). So yes, the *Jacobsen* court did find that the attribution and notice requirements of the Artistic License were conditions—but only after engaging in the necessary holistic contractual analysis in the context of determining whether a copyright registrant had the right to bring a copyright infringement claim. *See Ticketmaster*, 306 F. Supp. 3d at 1173 (determination of condition vs. covenant requires consideration of “the whole contract, its purpose, and the intention of the parties”).

Even as it cites *Jacobsen* with abandon in its brief, OpenAI makes some careful omissions which defeat its argument. Most conspicuously, OpenAI does not mention that *Jacobsen* dealt with whether an infringement claim could be brought, not whether a breach of contract claim could be brought under the Artistic License’s terms. Indeed, OpenAI does not cite (nor have Plaintiffs been able to identify) a single case dismissing a claim where a plaintiff was suing under an open-source license on the grounds that no contract was formed because the terms of the relevant license were conditions rather than covenants. Indeed, *none* of the cases OpenAI cites in support concern breach-of-contract claims but rather copyright-infringement claims.

There is good reason why no cases have dismissed contract claims based on OpenAI’s novel argument—it is black letter contract law that assent is manifested once intent to perform can be inferred. *See* Restatement (Second) of Contracts § 2 (1981) (“A promisor manifests an intention if he believes or has reason to believe that the promise will infer that intention from his words or his conduct”). *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996) is instructive. In *ProCD*, the court was confronted with a novel issue at the time—whether a license, which was

1 encoded on a compact disc, printed in a manual, and that appeared every time a software runs is  
2 binding on the user when the box containing the consumer product merely stated that there were  
3 restrictions contained in an enclosed license. *Id.* at 1450. Defendant bought the consumer package  
4 of the software but decided to ignore the license. *Id.* The court, applying Wisconsin law,  
5 concluded the license was binding and gave the example of a purchaser of an airline ticket:

6 Or consider the purchase of an airline ticket. The traveler calls the carrier  
7 or an agent, is quoted a price, reserves a seat, pays, and gets a ticket, in  
8 that order. The ticket contains elaborate terms, which the traveler can  
9 reject by canceling the reservation. To use the ticket is to accept the  
10 terms, even terms that in retrospect are disadvantageous.

11 *Id.* at 1451.

12 Those principles are no less applicable here. If OpenAI did not wish to be bound by the  
13 terms of the Suggested Licenses that adhered to using Plaintiffs' code, it simply could choose not  
14 to use them. Indeed, there is code that is published without any license at all. *See, e.g.*, ¶ 185.  
15 Rather, OpenAI chose to use Plaintiffs' code, thereby accepting the terms of the Suggested  
16 Licenses. OpenAI ignored the licenses instead. That is a breach of contract. Any other answer  
17 would undermine the entire software industry by rendering open-source programmers potentially  
18 liable for failures to perform while allowing users unfettered use of software without limitation.  
19 *See ProCD*, 86 F.3d at 1451-52 (explaining that not recognizing software licenses would render  
20 those transactions "unfettered by terms—so the seller has made a broad warranty and must pay  
21 consequential damages for any shortfalls in performance, two 'promises' that if taken seriously  
22 would drive prices through the ceiling or return transactions to the horse-and-buggy age." ).  
23 Whether a copyright infringement claim is available is immaterial. *Cf. SFC*, 2022 WL 1527518, at  
24 \*4 ("Potentially, there is an argument that Vizio's distribution of the software violates various the  
25 conditions of the software's copyright; however, SFC has not chosen to bring such claim. *And,*  
26 *indeed, because SFC is not the copyright holder, it cannot even assert one.*") (emphasis added).<sup>3</sup>

---

27 <sup>3</sup> Plaintiffs observe that unlike here, *Jacobsen* involved code that had been registered with the  
28 Copyright Office; hence, *Jacobsen* could seek relief under the Copyright Act through an  
infringement action. While Plaintiffs in this case maintain a copyright interest in their code, *see,*  
*e.g.*, SAC ¶¶ 19-23, none have registered their code with the Copyright Office—indeed, to do so  
would defeat the purpose of open source and most software code is not registered. OpenAI, itself,



1 Remarkably, OpenAI omits to mention that the Artistic License is **not** one of the OSLs at  
2 issue in this case. *See generally* OpenAI Br. at nowhere. The Artistic License is not one of the 11  
3 “Suggested Licenses” that form the basis of Plaintiffs’ claims. SAC at 10, n.4. So *Jacobsen*, even  
4 on its own terms, is merely advisory for this case.

5 Beyond that, none of the OSLs in this case—the aforementioned Suggested Licenses—  
6 contain same clear intent as the Artistic License at issue in *Jacobsen* to create conditions and not  
7 covenants. Although many of the OSLs at issue here include “provided that” language in sections  
8 regarding reproduction and redistribution, these provisions do not operate pursuant to a clear,  
9 overarching expressions of intent that the license will be granted only upon their satisfaction.<sup>4</sup> *See*  
10 OpenAI Br. at 10–17. In fact, other language is clearly to the contrary.

11 Although OpenAI has scoured the OSLs for conditional language—e.g., “provided that”  
12 and “conditions”—the mere fact that conditional language is used in a license does not  
13 necessarily indicate that a condition is intended. *Ticketmaster*, 306 F. Supp. 3d at 1173. To treat  
14 these terms as merely conditions with respect to the Suggested Licenses would undermine the  
15 most basic tenet of contract law, i.e., that terms in a contract are to be read holistically and in  
16 keeping with the parties’ intent. *Ticketmaster*, 306 F. Supp. 3d at 1173 (citation omitted). This is  
17 especially true in this case, where the Suggested Licenses were drafted by different hands, in  
18 different jurisdictions, at different times.

19 Further, because none of the OSLs in this case unambiguously suggest that the licensors  
20 intended the attribution and notice requirements to operate as conditions precedent to the grant  
21

---

22 has never registered a copyright. Instead, it typically publishes its open-source code through  
23 OSLs, including two used by Plaintiffs in this case (Apache and MIT). OpenAI’s position here  
24 would mean that software coders, and companies like OpenAI who routinely use OSLs, are out of  
25 luck to enforce the terms of their OSLs *unless* they have registered their code. If not, they cannot  
26 bring a breach of contract claim (because, as OpenAI argues, no license was granted), and they  
27 cannot bring a copyright infringement claim (because, as OpenAI contends, such claims require  
28 registration). To interpret the OSLs in this case as such—and OSLs more generally—would  
imperil the entire ecosystem of open-source licensing, including OpenAI’s.

<sup>4</sup> For instance, the Eclipse Public License 2.0 expressly opens with stating: “ANY USE,  
REPRODUCTION, OR DISTRIBUTION OF THE PROGRAM CONSTITUTES  
RECIPIENT’S ACCEPTANCE OF THIS AGREEMENT.” ECF No. 1-1, Appendix A at 11.



1 of the license, this Court should decline OpenAI’s invitation to treat the terms merely as  
2 conditions that strip the rights of software coders to enforce their licenses. *Netbula*, 2008 WL  
3 228036, at \*3.

4 **b. Plaintiffs’ allegations of OSL violations include “copyleft”**  
5 **requirements that OpenAI completely fails to address**

6 Moreover, *Jacobsen* pertained only to violations of attribution and notice requirements.  
7 *Jacobsen*, 609 F. Supp. 2d at 933. Here, however, Plaintiffs do not simply allege that OpenAI  
8 violated attribution and notice requirements. They also allege that OpenAI “fail[ed] to provide  
9 the source code of Copilot nor a written offer to provide the source code upon the request of each  
10 licensee,” SAC ¶ 254, and failed to provide a copy of the corresponding OSLs with any “copy,  
11 redistribution, or derivative of the software code,” *id.* at ¶ 184. In other words, Plaintiffs allege  
12 violations pertaining to other features of the OSLs that courts since *Jacobsen* have recognized  
13 properly fall under contract law. *Artifex Software*, 2017 WL 1477373, at \*3 (quoting *Versata*  
14 *Software, Inc. v. Ameriprise Fin., Inc.*, No. A-14-CA-12-SS, 2014 WL 950065, at \*5 (W.D. Tex.  
15 Mar. 11, 2014)).

16 In open-source licensing parlance, license provisions that require users to publish their  
17 modifications of licensed source code are called “copyleft” requirements. Copyleft requirements  
18 are not a feature of every open-source license, but they are part of six of the 11 Suggested  
19 Licenses at issue in this case.<sup>5</sup> These OSLs are also among the most ubiquitous OSLs and are  
20 commonly attached to code published in GitHub’s public repositories. For instance, the four  
21 GNU copyleft licenses open with a Preamble that celebrates “free software” and the right to  
22 “receive source code or . . . get it if you want it,” including the right to “change the software or  
23 use pieces of it in new free programs; *and that you know you can do these things.*” *See id.* at 16, 29,  
24 35, 49 (emphasis added).

25 Every court that has considered breach of copyleft provisions in an OSL has held that

---

26 <sup>5</sup> The Suggested Licenses are the Eclipse Public License 2, the GNU Affero General Public  
27 License version 3, the GNU General Public License version 2, the GNU General Public License  
28 version 3, the GNU Lesser General Public License version 2.1, and the Mozilla Public License  
2.0. *See* ECF No. 1-1, Appendix A.

such claims are the subject of contract law. *Artifex Software*, 2017 WL 1477373, at \*3; *Versata Software*, 2014 WL 950065, at \*5; *SFC*, 2022 WL 1527518, \* 3. In *Versata Software*, for example, the district court considered whether a counterclaim for breach of the GPL’s Source Code Provision was preempted by copyright law. The *Versata* court rejected the argument that this provision “amount[ed] to nothing more than a promise to not commit copyright infringement”—in other words, a right that was not recognized as one of the exclusive rights protected by copyright. *Versata Software*, 2014 WL 950065, at \*5. As the court explained, the plaintiff breached an additional obligation: “an affirmative promise to make its derivative work open source because it incorporated an open source program into its software,” which was “an additional contractual promise separate and distinct from any rights provided by the copyright laws.” *Id.*

Similarly, in *Artifex Software*, the district court held that the Copyright Act did not preempt a contract claim alleging a breach of the GPL’s requirement to share source code, recognizing the GPU GPL’s open-source requirement to be an “extra element” or additional contractual promise. *Artifex Software*, 2017 WL 1477373, at \*3. And more recently in *SFC*, 2022 WL 1527518, the district court squarely recognized that there was “no right to receive certain works—or source code in particular—under the Copyright Act” and that such a right would in fact appear to be “*the very opposite*” of those exclusive rights recognized under the Act. *Id.* at \*3. Thus, the court upheld a breach of contract claim in relation to the very same breach allegations Plaintiffs make here.

So, too, should the Court here. Plaintiffs have plausibly alleged that Copilot outputs source code licensed under OSLs that include copyleft provisions, and that the failure of Copilot to provide the source code or the corresponding license directly contravenes this scheme. SAC ¶¶ 13-20. Neither of these rights are protected by the Copyright Act. Therefore, even if the Court determined that the attribution and notice requirements in this case are conditions rather than covenants that limit the scope of the OSLs, Plaintiffs have still plausibly alleged breaches of copyleft provisions, which courts have routinely found to be covenants under contract law.

#### IV. CONCLUSION

For the foregoing reasons, OpenAI’s motion to dismiss the SAC should be denied.

1 Dated: March 27, 2024

By: /s/ Joseph R. Saveri

Joseph R. Saveri

Joseph R. Saveri (State Bar No. 130064)

Cadio Zirpoli (State Bar No. 179108)

Christopher K.L. Young (State Bar No. 318371)

Louis A. Kessler (State Bar No. 243703)

Elissa A. Buchanan (State Bar No. 249996)

William W. Castillo Guardado (State Bar No. 294159)

Holden J. Benon (State Bar No. 325847)

**JOSEPH SAVERI LAW FIRM, LLP**

601 California Street, Suite 1000

San Francisco, CA 94108

Telephone: (415) 500-6800

Facsimile: (415) 395-9940

Email: jsaveri@saverilawfirm.com

czirpoli@saverilawfirm.com

cyoung@saverilawfirm.com

lkessler@saverilawfirm.com

eabuchanan@saverilawfirm.com

wcastillo@saverilawfirm.com

hbenon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)

1920 Hillhurst Avenue, #406

Los Angeles, CA 90027

Telephone: (323) 968-2632

Facsimile: (415) 395-9940

Email: mb@buttericklaw.com

*Counsel for Individual and Representative  
Plaintiffs and the Proposed Class*



# EXHIBIT E

UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA

J. DOE 1, et al.,  
Plaintiffs,

v.

GITHUB, INC., et al.,  
Defendants.

Case No. 22-cv-06823-JST

**ORDER DENYING MOTION FOR  
RECONSIDERATION**

Re: ECF No. 218

Before the Court is Plaintiffs' motion for reconsideration, ECF No. 218, of the Court's prior order, ECF No. 189. The Court's prior order dismissed Plaintiffs' claims under Sections 1202(b)(1) and 1202(b)(3) of the DMCA with leave to amend. It also dismissed Plaintiffs' state law claims for intentional and negligent interference with prospective economic relations, unjust enrichment, negligence, and unfair competition with prejudice. *Id.* The Court will deny the motion for reconsideration.

The Court possesses the inherent authority to reconsider its interlocutory orders at any point before it enters final judgment. Fed. R. Civ. P. 54(b); *Amarel v. Connell*, 102 F.3d 1494, 1515 (9th Cir. 1996). Typically, a motion for reconsideration should be granted only if "the district court is presented with newly discovered evidence, committed clear error, or if there is an intervening change in the controlling law." *Kona Enterprises, Inc. v. Estate of Bishop*, 229 F.3d 877, 890 (9th Cir. 2000) (internal quotation marks and citations omitted).

In this district, a motion for leave to file a motion for reconsideration must comply with Civil Local Rule 7-9. Thus, the moving party must show reasonable diligence in bringing the motion and one of the following:

1. That at the time of the motion for leave, a material difference in

fact or law exists from that which was presented to the Court before entry of the interlocutory order for which reconsideration is sought. The party also must show that in the exercise of reasonable diligence the party applying for reconsideration did not know such fact or law at the time of the interlocutory order; or

2. The emergence of new material facts or a change of law occurring after the time of such order; or

3. A manifest failure by the Court to consider material facts or dispositive legal arguments which were presented to the Court before such interlocutory order.

Civ. L.R. 7-9(b).

First, Plaintiffs argue that “the DMCA injunctive claims for Does 3 and 4 should have remained intact, but instead they were dismissed.” ECF No. 218 at 5. The Court stands by its ruling. While the Court found that Plaintiffs had standing to seek injunctive relief, that did not necessarily entail that Plaintiffs stated a Section 1202 claim. As noted by Defendant GitHub, “[t]o say that a plaintiff has plausibly alleged an injury sufficient to seek injunctive relief under Article III is not the same thing as saying that a plaintiff has stated a claim for injunctive relief.” ECF No. 228 at 5. In its prior order, the Court found that Plaintiffs failed to allege that output from Copilot was “identical,” as their amended complaint alleged that the output is often “a modification.” See ECF No. 189 at 15 (quoting ECF No. 97-3 ¶ 96). Plaintiffs’ motion does not identify a material difference in law or fact, the emergence of new material facts or law, or even a manifest failure by the Court to consider material facts or dispositive legal arguments to support their position. See Civ. L.R. 7-9(b). Accordingly, Plaintiffs fail to satisfy the requirements for reconsideration for this argument.

Second, Plaintiffs argue that the the Court failed to consider the decision in *GC2 v. Int’l Game Tech.*, 391 F. Supp. 3d 828 (N.D. Ill. 2019), “despite its inclusion in Plaintiffs’ oppositions to Defendants’ first motions to dismiss.” ECF No. 218 at 6. As a threshold matter, Plaintiffs failed to include this case in their most recent opposition. The Court need not “hunt[]” for arguments “buried in briefs” filed long ago. *Doyle v. Galderma, Inc.*, No. 19-cv05678-TSH, 2021 WL 1721069, at \*14 (N.D. Cal. Apr. 30, 2021). But regardless, *GC2* is inapposite as it concerned the removal of CMI from artwork that was copied “in its entirety.” 391 F. Supp. 3d at 843–44. In their amended complaint, Plaintiffs only alleged that snippets of code were reproduced, which is distinguishable. Thus, Plaintiffs’ argument does not meet the standard for reconsideration.



1 Finally, Plaintiffs contend that “the Court overlooked where Plaintiffs specifically alleged  
2 identical output of Doe 5’s Licensed Materials.” ECF No. 218 at 7. Not so. The Court  
3 considered Doe 5’s allegations and rejected them. *See* ECF No. 189 at 3–4, 15 (“[T]he examples  
4 Plaintiffs provide with respect to Does 1, 2, and 5 state that the Copilot output is a ‘modified  
5 format,’ ‘variation[],’ or the ‘functional[] equivalent’ of the licensed code.”). Throughout the  
6 amended complaint, there was no showing of identical output of any entire work by Doe 5, and  
7 Plaintiffs point to nothing in their present motion that warrants reconsideration of that conclusion.

8 For the foregoing reasons, Plaintiffs’ motion for reconsideration regarding the Court’s  
9 prior motion to dismiss order is denied.

10 **IT IS SO ORDERED.**

11 Dated: April 15, 2024

12   
13 JON S. TIGAR  
14 United States District Judge  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

# **EXHIBIT F**

# Exhibit A



JOSEPH SAVERI  
LAW FIRM

601 CALIFORNIA STREET  
SUITE 1505  
SAN FRANCISCO CA 94108

TEL 415.500.6800  
FAX 415.395.9940

June 25, 2024

**Via Electronic Mail**

Joseph Gratz  
Morrison & Foerster LLP  
425 Market Street  
San Francisco, CA 94105-2482

**Re: *Doe 1 v. GitHub, Inc*, Lead Case No. 4:22-cv-06823-JST**

Dear Joe:

Plaintiffs are aware that OpenAI produced documents in the *Tremblay v. OpenAI* matter. Many of those documents appear to be relevant to this case. To the extent that these documents are responsive to Plaintiffs' RFPs, Plaintiffs request that OpenAI produce them given that there is no burden in producing them given that they have already been collected and that the proportionality considerations set forth in Rule 26 support the discovery.

Plaintiffs also note that Judge Tigar's narrowing of the claims in this case does not alter the relevance of the discovery sought by Plaintiffs. Many (if not all) of the discovery requests relevant to the DMCA claims and the breach of contract claim are co-extensive (*e.g.*, the removal of CMI for training purposes and the decimation of licensed code with CMI removed underscored Plaintiffs' Section 1202 claim, but this is also, simultaneously, the conduct underscoring our allegations that Defendants breached Plaintiffs' open-source licenses). Plaintiffs are reviewing their discovery requests to determine if any were solely relevant to the DMCA claims, but in the meantime believe the parties should continue to meet and confer to move discovery along.

To that end we write to follow up on promises made by OpenAI to produce relevant documents and supplement its responses to interrogatories, as detailed below.